

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Fri Aug 6 00:59:57 1999; MasPar time 1818.19 Seconds
Tabular output not generated. 1524.539 Million cell updates/sec

Title: >US-08-287-669-18
Description: (901-1900) from US08287669.seq (2 of 10)
Perfect Score: 1000
N.A. Sequence: 901 GTCGTCCTGGTATCCTCAA.....TGAAATTTGTTTATGTT 1900
Comp: CAGCAGGAACCATAGGAGTT.....ACTTTAACACAAATACAA

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl58
1:em_bal 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
7:em_in 8:em_om 9:em_or 10:em_ov 11:em_pat 12:em_ph
13:em_pl 14:em_ro 15:em_sts 16:em_vi
genbank111
Database: 17:gb_bal 18:gb_ba2 19:gb_htg1 20:gb_htg2 21:gb_in1
22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
28:gb_pl2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_vi

Statistics: Mean 11.164; Variance 7.504; scale 1.488

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description	Pred. No.
1	1000	100.0	7653	21	CELCED3A	Caenorhabditis elegans	0.00e+00
C 2	931	93.1	39908	21	CEC48D1	Caenorhabditis elegans	0.00e+00
C 3	925	92.5	314495	19	CEY67H2	Caenorhabditis elegans	0.00e+00
4	168	16.8	39908	21	CEC48D1	Caenorhabditis elegans	1.50e-78
C 5	163	16.3	207370	19	AC006798	Caenorhabditis elegans	1.59e-75
6	162	16.2	314495	19	CEY67H2	Caenorhabditis elegans	6.41e-75
7	160	16.0	306131	19	AC006874	Caenorhabditis elegans	1.04e-73
8	156	15.6	207370	19	AC006798	Caenorhabditis elegans	2.68e-71
C 9	156	15.6	299202	19	AC006792	Caenorhabditis elegans	2.68e-71
10	154	15.4	306131	19	AC006874	Caenorhabditis elegans	4.30e-70
C 11	148	14.8	32209	19	AC006663	Caenorhabditis elegans	1.75e-66
12	146	14.6	140702	20	AC006888	Caenorhabditis elegans	2.77e-65
13	146	14.6	317781	20	AC006906	Caenorhabditis elegans	2.77e-65

C	14	14.5	7653	21	CELCED3A	Caenorhabditis elegans	1.10e-64
15	144	14.4	125590	21	CEY49E10	Caenorhabditis elegans	4.38e-64
C 16	139	13.9	16938	22	CELT25D3	Caenorhabditis elegans	4.30e-61
C 17	139	13.9	125590	21	CEY49E10	Caenorhabditis elegans	4.30e-61
C 18	139	13.9	274498	19	AC006765	Caenorhabditis elegans	4.30e-61
19	138	13.8	143092	21	CEY17G7B	Caenorhabditis elegans	1.70e-60
20	138	13.8	207139	19	CEY17G7	Caenorhabditis elegans	1.70e-60
21	137	13.7	39752	22	CELC04F5	Caenorhabditis elegans	6.73e-60
22	137	13.7	206217	19	AC006754	Caenorhabditis elegans	6.73e-60
23	136	13.6	42724	21	CEP53A2	Caenorhabditis elegans	2.66e-59
24	134	13.4	297866	19	AC006794	Caenorhabditis elegans	4.14e-58
25	134	13.4	299202	19	AC006792	Caenorhabditis elegans	1.63e-57
C 26	133	13.3	262336	19	AC006786	Caenorhabditis elegans	6.43e-57
C 27	132	13.2	140702	20	AC006888	Caenorhabditis elegans	6.43e-57
C 28	132	13.2	143092	21	CEY17G7B	Caenorhabditis elegans	6.43e-57
C 29	132	13.2	207139	19	CEY17G7	Caenorhabditis elegans	6.43e-57
C 30	131	13.1	267118	20	AC006889	Caenorhabditis elegans	2.53e-56
C 31	129	12.9	254877	20	AC006904	Caenorhabditis elegans	3.90e-55
32	129	12.9	292390	19	AC006900	Caenorhabditis elegans	3.90e-55
C 33	128	12.8	32209	19	AC006663	Caenorhabditis elegans	1.53e-54
C 34	126	12.6	299727	19	AC006738	Caenorhabditis elegans	2.35e-53
35	125	12.5	19952	22	CELY38C9A	Caenorhabditis elegans	9.19e-53
C 36	125	12.5	38643	21	CELT03F1	Caenorhabditis elegans	9.19e-53
C 37	125	12.5	40662	21	CEY32B12B	Caenorhabditis elegans	9.19e-53
C 38	125	12.5	191857	19	CEY32B12	Caenorhabditis elegans	9.19e-53
C 39	125	12.5	299782	19	AC006844	Caenorhabditis elegans	3.59e-52
40	124	12.4	40662	21	CEY32B12B	Caenorhabditis elegans	3.59e-52
41	124	12.4	191857	19	CEY32B12	Caenorhabditis elegans	1.40e-51
42	123	12.3	35776	21	CELW02G9	Caenorhabditis elegans	5.47e-51
43	122	12.2	95968	21	CEY47D3B	Caenorhabditis elegans	5.47e-51
44	122	12.2	197735	19	CEY66A7	Caenorhabditis elegans	5.47e-51
45	122	12.2	298195	19	AC006890	Caenorhabditis elegans	5.47e-51

ALIGNMENTS

RESULT	1	CELCED3A	7653 bp	DNA	INV	23-FEB-1994
LOCUS		Caenorhabditis elegans cell death protein (ced-3) gene, complete cds.				
DEFINITION		L29052				
ACCESSION		g456416				
NID		L29052.1	GI:456416			
VERSION		cell death protein; interleukin-1 beta converting enzyme.				
KEYWORDS		Caenorhabditis elegans (strain N2) DNA.				
SOURCE		Caenorhabditis elegans				
ORGANISM		Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Caenorhabditis.				
REFERENCE		1 (bases 1 to 7653)				
AUTHORS		Yuan, J., Shaham, S., Ledoux, S., Ellis, H.M. and Horvitz, H.				
TITLE		The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1b-converting enzyme				
JOURNAL		Cell 75, 641-652 (1993)				
MEDLINE		94061982				
FEATURES		Location/Qualifiers				
Source		1..7653				
		/organism="Caenorhabditis elegans"				
		/strain="N2"				
exon		/db_xref="taxon:6239"				
		2167..2366				
		/gene="ced-3"				
gene		/number=1				
		join(2232..2366,2430..2575,2854..3107,4303..4634,5547..5760,5815..5942,6298..6537,7013..7075)				
		/gene="ced-3"				
CDS		join(2232..2366,2430..2575,2854..3107,4303..4634,5547..5760,5815..5942,6298..6537,7013..7075)				
		/gene="ced-3"				
		/note="codes for a protein similar to mammalian interleukin-1b-converting enzyme"				
		/codon_start=1				
		/product="cell death protein"				

/protein_id="AAA27982.1"
/db_xref="PID:9456417"
/db_xref="GI:456417"
/translation="MMQRDRRLRNIMFSSHLKVDLEILVLIKQVLSNDGDMI
NSCGTVREKRREIVKAVQRRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDSNAVEF
ECPMSPAHSRRSRLSPAGYTSPTRVHRDSSVSSVSTSYQDIYSRARSRSRALHS
FDHNTYSSPPVNAFPSPSSANSFTGCSLGSYSSSRNRSFSKASGPTQYIFHEEDMN
FVDAPTSRVFDEKTMYNFSSPRGMCLIIINNEHFEQMPTRNGTKADKDLNLFRCM
GYTVICKDNLGTGRMLLTIRDEAKHSHGDSAILVILSHGEENVLIIGVDDIPISHEI
YDLLNANAPRLANKPKIVFQACRGERRDNGFPVLDSDGVPFALRRGWDNRDGPLF
NFLGCVRPQVQVWRKKPSQADILIRYATTAQYVSWRNSARGSWF IQAVCEVFETHAK
MDVVELLETVNKKVACGFTSQGSNLIKQMPENTSRLLKKFYFWPEARNSAV"
2367. .2429
/gene="ced-3"
/number=1
2430. .2575
/gene="ced-3"
/number=2
2576. .2853
/gene="ced-3"
/number=2
2854. .3107
/gene="ced-3"
/number=3
3108. .4302
/gene="ced-3"
/number=3
4303. .4634
/gene="ced-3"
/number=4
4635. .5546
/gene="ced-3"
/number=4
5547. .5760
/gene="ced-3"
/number=5
5761. .5814
/gene="ced-3"
/number=5
5815. .5942
/gene="ced-3"
/number=6
5943. .6297
/gene="ced-3"
/number=6
6298. .6537
/gene="ced-3"
/number=7
6538. .7012
/gene="ced-3"
/number=7
7013. .7652
/gene="ced-3"
/number=8

intron

exon

intron

exon

intron

exon

intron

exon

intron

exon

intron

exon

intron

exon

BASE COUNT 2429 a 1455 c 1271 g 2498 t
ORIGIN

Query Match 100.0%; Score 1000; DB 21; Length 7653;
Best Local Similarity 100.0%; Pred. NO. 0.00e+00;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 901 GTCGTCCTTGGTATCCTCAACTGTCCCGGTTTTGTTTTTCGGTACACTTCTCCGTGATGC 960
|||||
QY 901 GTCGTCCTTGGTATCCTCAACTGTCCCGGTTTTGTTTTTCGGTACACTTCTCCGTGATGC 960
|||||
Db 961 CACCTGTCTCCGTCTCAATATATCGTTTAGAAATGTGAAGTGTCCAGATGGGTGACTCATA 1020
|||||
QY 961 CACCTGTCTCCGTCTCAATATATCGTTTAGAAATGTGAAGTGTCCAGATGGGTGACTCATA 1020
|||||
Db 1021 TTGCTGCTGCTACAATCCACCTTTCTTTTCTCATCGGCAGTCTTACGAGCCCATATAAAC 1080
|||||
QY 1021 TTGCTGCTGCTACAATCCACCTTTCTTTTCTCATCGGCAGTCTTACGAGCCCATATAAAC 1080
|||||

Db 1081 TTTTTCCTCCGCGAAATTTGCAATAAACCAGCGCAAAACCTTTCTCAAAATGTTACGCAA 1140
|||||
QY 1081 TTTTTCCTCCGCGAAATTTGCAATAAACCAGCGCAAAACCTTTCTCAAAATGTTACGCAA 1140
|||||
Db 1141 TATATACAAATCCATAAGAATATCTTCTCAATGTTTATGATTTCTTCGACGACTTTCTCT 1200
|||||
QY 1141 TATATACAAATCCATAAGAATATCTTCTCAATGTTTATGATTTCTTCGACGACTTTCTCT 1200
|||||
Db 1201 TCCTGTGCTAACATCTTATTTTATAATATTTCCGCTAAATAATCCGATTTTTCAGTATTA 1260
|||||
QY 1201 TCCTGTGCTAACATCTTATTTTATAATATTTCCGCTAAATAATCCGATTTTTCAGTATTA 1260
|||||
Db 1261 ATTTATCGTAAATAATATCATATAAGCACCGCAAAACTACTAAATAATGGTAAAGCTCCTTT 1320
|||||
QY 1261 ATTTATCGTAAATAATATCATATAAGCACCGCAAAACTACTAAATAATGGTAAAGCTCCTTT 1320
|||||
Db 1321 TAAATCGGCTCGACATTTATCGTATTAAGGAATCACAATAATCTGAGAATGCGTACTGCGC 1380
|||||
QY 1321 TAAATCGGCTCGACATTTATCGTATTAAGGAATCACAATAATCTGAGAATGCGTACTGCGC 1380
|||||
Db 1381 AACATATTTGACGGCAAAATATCTCGTAGCGAAACACAGTAATTTCTTTAAATGACTAC 1440
|||||
QY 1381 AACATATTTGACGGCAAAATATCTCGTAGCGAAACACAGTAATTTCTTTAAATGACTAC 1440
|||||
Db 1441 TGTAGCGCTTGTGTCGATTTACGGGCTCAATTTTGAATAATAATTTTTTTTCGAAATTT 1500
|||||
QY 1441 TGTAGCGCTTGTGTCGATTTACGGGCTCAATTTTGAATAATAATTTTTTTTCGAAATTT 1500
|||||
Db 1501 TGATAACCCGTAATAATCGTCACACGCTACAGTAGTCAATTAAGGATTTACTGTAGTTCTA 1560
|||||
QY 1501 TGATAACCCGTAATAATCGTCACACGCTACAGTAGTCAATTAAGGATTTACTGTAGTTCTA 1560
|||||
Db 1561 GCTACGAGATATTTTGGCGGCCCAATATGACTGTAAATACGCATTTCTGAATTTTGTGTT 1620
|||||
QY 1561 GCTACGAGATATTTTGGCGGCCCAATATGACTGTAAATACGCATTTCTGAATTTTGTGTT 1620
|||||
Db 1621 TCCGTAATAATTTTCAACAAGATTTTGGCATTTCCACTTTAAAGCGCACAGGATTTATTTCCA 1680
|||||
QY 1621 TCCGTAATAATTTTCAACAAGATTTTGGCATTTCCACTTTAAAGCGCACAGGATTTATTTCCA 1680
|||||
Db 1681 ATGGGTCTCGGCACGCAAAAGTTTGTAGTACACTTTTAAATTTCTCCTTGCATTTTAAATTC 1740
|||||
QY 1681 ATGGGTCTCGGCACGCAAAAGTTTGTAGTACACTTTTAAATTTCTCCTTGCATTTTAAATTC 1740
|||||
Db 1741 AATTACTAAAAATTTTTCGTGAATTTTCTGTAAAAATTTTAAATTCAGTTTCTTAATATTT 1800
|||||
QY 1741 AATTACTAAAAATTTTTCGTGAATTTTCTGTAAAAATTTTAAATTCAGTTTCTTAATATTT 1800
|||||
Db 1801 TTCCAGGCTGACAAACAGAAACACAAACACAAACACAAACATTTTAAATTCAGTTTCTTAAT 1860
|||||
QY 1801 TTCCAGGCTGACAAACAGAAACACAAACACAAACACAAACATTTTAAATTCAGTTTCTTAAT 1860
|||||
Db 1861 TAAAAATAACGATTTTCTCATTTGAAAAATTTGTTTTATGTT 1900
|||||
QY 1861 TAAAAATAACGATTTTCTCATTTGAAAAATTTGTTTTATGTT 1900
|||||

RESULT 2
LOCUS CEC48D1 39908 bp DNA INV 23-NOV-1998
DEFINITION Caenorhabditis elegans cosmid C48D1, complete sequence.
ACCESSION 281049
NID g1627677
VERSION 281049.1 GI:1627677
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 39908)
AUTHORS Burton, J.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rw@nematode.wustl.edu

CP 1660 TTTAAAGTGAATGCCAAATCTTGTGAAATTTATACGGAACACAAAATTCAGAGAATG 1601

Db 11754 CGTATTACGCATCATATTTGGCGCGCAAAATATCTCGTAGCTAGAACTACAGTAATCCIT 11813

CP 1600 CGTATTACAG-TCATATTTGGCGCGCAAAATATCTCGTAGCTAGAACTACAGTAATCCIT 1542

Db 11814 TAAATGACTACCGTAGCGCTTGTGACGATTTACGGGTTGTCAAAATTCGAAAAAATAAT 11873

CP 1541 TAAATGACTACTGTAGCG-TTGTGACGATTTACGGGTTATCAAAATTCGAAAAAATAAT 1483

Db 11874 TATTTTCAAAAATTTAGCCCGTAATCGACACAAAGCGCTACAGTAGTCATTTAAAGAATT 11933

CP 1482 TATTTTCAAAAATTTAGCCCGTAATCGACACAAAGCGCTACAGTAGTCATTTAAAGAATT 1423

Db 11934 ACTGTAGTTTTCGCTACGAGATATTTTTCGCGGTCAAAATATGTTGGCAGTACGCAATCTC 11993

CP 1422 ACTGTAGTTTTCGCTACGAGATATTTTTCG-CGTCAAAATATGTTGGCAGTACGCAATCTC 1364

Db 11994 AGAATTTTGTGATTCCTTAATACGATAATGTCGAGCGGATTTAAAG-A-CTTT-ACCAT 12050

CP 1363 AGAATTTTGTGATTCCTTAATACGATAATGTCGAGCGGATTTAAAGGAGCTTTTACCAT 1304

Db 12051 TTTTGTAGTTTTCGCTGCTATTATGATAATTTTACGATAATTAATACTCAAAATCGG 12110

CP 1303 TTTTGTAGTTTTCGCTGCTATTATGATAATTTTACGATAATTAATACTCAAAATCGG 1244

Db 12111 AATTTAGCGGAATATATATAAAATAAGATGTTAGCACGAGAGAAAGTGCTGCGAA 12170

CP 1243 AATTTAGCGGAATATATATAAAATAAGATGTTAGCACGAGAGAAAGTGCTGCGAA 1184

Db 12171 GAAATCATAAACATTTGAGAAGATATTTCTTATGATTTGATTTGATTTGATTTGGA 12230

CP 1183 GAAATCATAAACATTTGAGAAGATATTTCTTATGATTTGATTTGATTTGATTTGGA 1124

Db 12231 GAAAGTTTTCGCGGTTTATTTGCAAAATTTTCGCGGAAAAAAGTTTATGATGGGCTCGT 12290

CP 1123 GAAAGTTTTCGCGGTTTATTTGCAAAATTTTCGCGGAAAAAAGTTTATGATGGGCTCGT 1064

Db 12291 AAGCATGCCGATGAGAAAGAAAGTGGATTGTAGCAGCAGCAATATGAGTCACCCATCTG 12350

CP 1063 AAGACTGCCGATGAGAAAGAAAGTGGATTGTAGCAGCAGCAATATGAGTCACCCATCTG 1004

Db 12351 GACAGTTTACATTTCTAAACGATAATTTGAGACGGAG-CAGGTGGCATCAGGAAGAGTGT 12409

CP 1003 GACAGTTTACATTTCTAAACGATAATTTGAGACGGAGCAGGTGGCATCAGGAAGAGTGT 944

Db 12410 ACCGAAACAAAACCGGGACAAAGTTGAGGATACCAAGGACGAC 12452

CP 943 ACCGAAACAAAACCGGGACAAAGTTGAGGATACCAAGGACGAC 901

RESULT 3 CEY67H2 314495 bp DNA HTG 04-MAR-1999

LOCUS Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from

INITIATION Clone Y67H2, WORKING DRAFT SEQUENCE.

ACCESSION AL022475

NID 94469034

VERSION AL022475.3 GI:4469034

KEYWORDS HTG; HTGS_PHASE1

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans

REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;

AUTHORS Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

TITLE 1 (bases 1 to 314495)

JOURNAL McMurray, A.

Submitted (04-MAR-1999) Nematode Sequencing Project, Sanger Centre,

Hinxton, Cambridge CB10 1RO, UK and Department of Genetics,

Washington University, St. Louis, MO 63110, USA. E-mail:

jes@sanger.ac.uk or rwenematode.wustl.edu

On Mar 22, 1999 this sequence version replaced gi:4468145.

IMPORTANT: This sequence is unfinished and does not necessarily

represent the correct sequence. Work on the sequence is in progress

and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments.

* NOTE: This is a 'working draft' sequence.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

FEATURES

Location/Qualifiers

1..314495

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="IV"

/clone="Y67H2"

BASE COUNT 96542 a 54551 c 52760 g 97027 t 13615 others

ORIGIN

Query Match 92.5%; Score 925; DB 19; Length 314495;

Best Local Similarity 98.3%; Pred. No. 0.00e+00;

Matches 987; Conservative 0; Mismatches 8; Indels 9; Gaps 9;

Db 241112 AACATAAAACACAAATTTCAATGAGAAATCGTTATTTTAAATTTGAAAACGATTTTAA 241171

CP 1900 AACATAAAACACAAATTTCAATGAGAAATCGTTATTTTAAATTTGAAAACGATTTTAA 1841

Db 241172 AATGTTTGTGTTTGTGTTTCTGTTGTCAGCCCTGGAAATATTAGAAAACGATTTT 241231

CP 1840 AATGTTTGTGTTTGTGTTTCTGTTGTCAGCCCTGGAAATATTAGAAAACGATTTT 1781

Db 241232 AAAAAATTTTACAGAAAAATTTACGAAAAATTTTAGTAATGAATTTAAAAATGCAAGGAGA 241291

CP 1780 AAAAAATTTTACAGAAAAATTTACGAAAAATTTTAGTAATGAATTTAAAAATGCAAGGAGA 1721

Db 241292 ATTTAAAGTCTATCAAACTTTTTCGCTGCCGAGACCCATTGGAATAAATCC-GTGCGCC 241350

CP 1720 ATTTAAAGTCTATCAAACTTTTTCGCTGCCGAGACCCATTGGAATAAATCCGTGCGCC 1661

Db 241351 TTTAAAGAGGAATGCCAAAATCTTGTGAAATATTACGAAAAACACAAAATTCAGAGAATG 241410

CP 1660 TTTAAAGTGAATGCCAAAATCTTGTGAAATATTACGAAAAACACAAAATTCAGAGAATG 1601

Db 241411 CGTATTACGCATCATATTTGGCGCGCAAAATATCTCGTAGCTAGAACTACAGTACATCCT 241470

CP 1600 CGTATTACAG-TCATATTTGGCGCGCAAAATATCTCGTAGCTAGAACTACAGTA-ATCCT 1543

Db 241471 TTAATGACTACCGTAGCGCTTGTGACGATTTACGSGTTGTCAAAATTCGAAAAAATAAT 241530

CP 1542 TTAATGACTACTGTAGCG-TTGTGACGATTTACGSGTTATCAAAATTCGAAAAAATAAT 1484

Db 241531 TTATTTTCAAAAATTTGAGCCCGTAAATCGACACAAAGCGCTACAGTAGTCATTTAAAGAAT 241590

CP 1483 TTATTTTCAAAAATTTGAGCCCGTAAATCGACACAAAGCGCTACAGTAGTCATTTAAAGAAT 1424

Db 241591 TACTGTAGTTTTCGCTACGAGATATTTTTCGCGCTCAAAATATGTTGCGCAGTACGCAATCT 241650

CP 1423 TACTGTAGTTTTCGCTACGAGATATTTTTCG-CGTCAAAATATGTTGCGCAGTACGCAATCT 1365

Db 241651 CAGAAATTTGTGATTCCTTAATACGATAATGTCGAGCCGATTTTAAAG-A-CTTT-ACCA 241707

CP 1364 CAGAAATTTGTGATTCCTTAATACGATAATGTCGAGCCGATTTTAAAGGAGCTTTTACCA 1305

Db 241708 TTTTGTAGTTTTCGCTGCTATTATGATAATTTTACGATAAATTAATACTCAAAAATCG 241767

CP 1304 TTTTGTAGTTTTCGCTGCTATTATGATAATTTTACGATAAATTAATACTCAAAAATCG 1245

Db 241768 GAATTTAGCGGAATATATATAAAATAAGATGTTAGCACACGAGAGAAAGTGCTGCGA 241827

CP 1244 GAATTTAGCGGAATATATATAAAATAAGATGTTAGCACACGAGAGAAAGTGCTGCGA 1185

Db 241828 AGAAATCATAAACATTTGAGAAGATATTTCTTATGATTTGATTTGATTTGCGTAACAATTTGG 241887

CP 1184 AGAAATCATAAACATTTGAGAAGATATTTCTTATGATTTGATTTGATTTGCGTAACAATTTGG 1125

Db 241888 AGAAAGTTTTGGCCGGTTTATTGCAAAATTCGCGGAAAAAAGTTTATGATGGGCTCG 241947
|||||
Cp 1124 AGAAAGTTTTGGCCGGTTTATTGCAAAATTCGCGGAAAAAAGTTTATGATGGGCTCG 1065
|||||
Db 241948 TAAGCATGCCGATGAGAAAAAGAAAGTGGATTGTAGCAGCAGCAATATGAGTCACCCATCT 242007
|||||
Cp 1064 TAAGACTGCCGATGAGAAAAAGAAAGTGGATTGTAGCAGCAGCAATATGAGTCACCCATCT 1005
|||||
Db 242008 GGACAGTTCACATTTCTAAACGATAATTGAGACGGAG-CAGGTGGATCAGCGAAGAGTG 242066
|||||
Cp 1004 GGACAGTTCACATTTCTAAACGATAATTGAGACGGAGACAGGTGGCATCACGGAAGAGTG 945
|||||
Db 242067 TACCGAAAAACAAACCGGACAAAGTTGAGGATFACCAAGGACGAC 242110
|||||
Cp 944 TACCGAAAAACAAACCGGACAAAGTTGAGGATFACCAAGGACGAC 901
|||||

4 CEC48D1 39908 bp DNA INV 23-NOV-1998
Caenorhabditis elegans cosmid C48D1, complete sequence.
281049
91627677
281049.1 GI:1627677
HTG.
Caenorhabditis elegans.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 39908)
Burton, J.
Direct Submission
Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rw@nematode.wustl.edu
2 (bases 1 to 39908)
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roope, A.,
Saunders, D., Shownkeen, R., Smalton, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Wohldman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans

JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
COMMENT Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.
For a graphical representation of this sequence and its analysis see:-
http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence&object=C48D1
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
This sequence is the entire insert of clone C48D1. The true right end of clone F58D2 is at 18510 in this sequence. The start of this sequence (1..100) overlaps with the end of sequence Z81093. The end of this sequence (33020..39908) overlaps with the start of sequence Z82274.
FEATURES
Location/Qualifiers
1..39908

gene
CDS
gene
CDS
gene
CDS

/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="C48D1"
complement(994..4439)
/gene="C48D1.1"
complement(join(994..1020,1345..1951,2283..2580,2629..2765,2882..3147,3615..3739,4089..4331,4382..4439))
/note="similar to C.elegans ZYG-11 like protein; cDNA EST EMBL:T01768 comes from this gene"
/codon_start=1
/protein_id="CAB02846.1"
/db_xref="PID:e1344741"
/db_xref="PID:g3875048"
/db_xref="GI:3875048"
/db_xref="SPTREMBL:O02227"
/translation="MTKPLHDIACOKIAEFVQNGSYDNRFEELDSKSSNIVSYLMNRDKFENLKTTHKLVNTKVTLPQTLDELTARQLKQFSLKELKIGNFSKSKSWCIFRRKE TRHVDIVRILEAVLNEETRNKQLYHIEKDQKLMRWTAKTGNMPLSLTHTLMSDVHI NEDDFYQLFCNFSNLRSLDISGNTNINAGISRLKQLEVLALRDFSIATYTELKDLFN LKHLRVLDVSQTRFNLRHSNNIVEQFECKMELPELREMDCTRTSLSGDILECLLKTQ LSIPNIIISVTEAETSQYNDVRVLTAYFHKSVQALVYCLTARKEFDSGYCMKAIRISI SEDNLQNLNIPSCIKTIKVSQMTFDSISVTQNGKLKLIIFGMLYSNQLCSNGVTLLVE ALMSFVLYKRMISEAYKVNISYWNAVEKLVNKLKLEDFKISKTAMSFLENRVSKALRTRASSVLCQCIHRMHSDELKSLSEKTEVNRLVDILEWSVGQKNETKLSEFFEITIKMTEKSSKACTNMIIQAGIISISLSCMEKVESNDITTKCLELTLNLSALNYDSLQELYTSSNVDKIGQANAYSTEK"
complement(6258..11120)
/gene="C48D1.2"
complement(join(6258..6349,6876..7065,7420..7547,7602..7815,8728..9059,10248..10501,10780..10925,10989..11120))
/note="similar to Caspase recruitment domain, ICE-like protease (caspase) p20 domain., ICE-like protease (caspase) p10 domain.; cDNA EST EMBL:D75533 comes from this gene; cDNA EST yk198f10.5 comes from this gene; cDNA EST yk491a5.5 comes from this gene"
/codon_start=1
/protein_id="CAB02848.1"
/db_xref="PID:e1344743"
/db_xref="PID:g3875050"
/db_xref="GI:3875050"
/db_xref="SPTREMBL:O02229"
/translation="MRQDRSLLERNIMFSSHLKVDEILEVLIQVLSNDGMDIN SCGTVREKRRREIVKAVQRGDAFADFALRSTGHEGLAEVLEPLARSVDSNAVEFE CPMSPASHRRSRALSPAGYTSPTRVHRDSVSSVSTSIQDIYSRARSRSRSLHSS DRHNYSSPPVNAFPSPSSANSSTGCSLGSYSSRNRSFSKASGPTQYIFHEEDMNE VDAPTISRVEDEKTYMRFNFSPPRGMCCLIINNEHFEQMPTRNGTKADKDLNLTNLFRCMG YTVICKDNLITGRGMLTIRDFAKHSHGDSAILVLSHGEENVIIIGVDDIPISTHEIY DLLNAANAPRLANKPKIVFQACRGERRDNGFPVLDSDVGVPAPFLRWGNDRDGPLEN FLGCVRPQVQVWRKPSQADILIAAYATTAQVSVWRNSARGSWFIQAVCEVFSTHAKD MDVVELLTEVNKKVAYDIPPAQKVLLLAGSTKLCRLKFTTRDSLPN"
complement(16940..21616)
/gene="C48D1.3"
complement(join(16940..17126,18419..18597,18985..19442,19553..19879,19925..19961,20333..20481,20789..20987,21037..21152,21373..21616))
/note="Similarity to Salmonella sodium/proline symporter (SW:PUTP_SALTY); cDNA EST EMBL:C13840 comes from this gene; cDNA EST EMBL:C11562 comes from this gene"
/codon_start=1
/protein_id="CAB02847.1"
/db_xref="PID:e1344742"
/db_xref="PID:g3875049"
/db_xref="GI:3875049"
/db_xref="SPTREMBL:O02228"
/translation="MADLLGIVAIIVFFVYLIVVGVWAGRKSSKSKESESEAGAATEE VMLAGRNIGTLVGIFTMTATVVGAYINGTAEALYNGLLGCQAPVGYAISLYMGGLL FAKKMRREGYITMLDPFPQFWNFLELIFGRTFDNFRKLGRFLKLTQIIEILDFFQHYG

100


```

8
RESULT
LOCUS      AC006798      207370 bp      DNA
DEFINITION Caenorhabditis elegans clone Y51F8, WORKING DRAFT SEQUENCE, 30
unordered pieces.
ACCESSION  AC006798
NID        94309785
VERSION    AC006798.2 GI:4309785
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     Caenorhabditis elegans.
ORGANISM   Caenorhabditis elegans.
REFERENCE  Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
AUTHORS    Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
TITLE      1 (bases 1 to 207370)
JOURNAL    Waterston, R.H.
REFERENCE  The sequence of Caenorhabditis elegans clone
AUTHORS    Unpublished
TITLE      2 (bases 1 to 207370)
JOURNAL    Waterston, R.H.
REFERENCE  Direct Submission
AUTHORS    Submitted (23-FEB-1999) Genome Sequencing Center, Washington
TITLE      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL    MO 63108, USA
COMMENT    On Mar 1, 1999 this sequence version replaced gi:4263140.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1986: contig of 1986 bp in length
* 1987 2000: gap of unknown length
* 2001 4562: contig of 2562 bp in length
* 4563 4576: gap of unknown length
* 4577 7877: contig of 3301 bp in length
* 7878 7891: gap of unknown length
* 7892 10172: contig of 2281 bp in length
* 10173 10186: gap of unknown length
* 10187 12395: contig of 2209 bp in length
* 12396 12409: gap of unknown length
* 12410 15306: contig of 2897 bp in length
* 15307 15320: gap of unknown length
* 15321 17774: contig of 2454 bp in length
* 17775 17788: gap of unknown length
* 17789 19933: contig of 2145 bp in length
* 19934 19947: gap of unknown length
* 19948 24444: contig of 4497 bp in length
* 24445 24458: gap of unknown length
* 24459 27765: contig of 3307 bp in length
* 27766 27779: gap of unknown length
* 27780 31869: contig of 4090 bp in length
* 31870 31883: gap of unknown length
* 31884 37836: contig of 5953 bp in length
* 37837 37850: gap of unknown length
* 37851 42797: contig of 4947 bp in length
* 42798 42811: gap of unknown length
* 42812 46266: contig of 3455 bp in length
* 46267 46280: gap of unknown length
* 46281 49619: contig of 3339 bp in length
* 49620 49633: gap of unknown length
* 49634 52707: contig of 3074 bp in length
* 52708 52721: gap of unknown length
* 52722 56266: contig of 3545 bp in length
* 56267 56280: gap of unknown length
* 56281 59691: contig of 3411 bp in length
* 59692 59705: gap of unknown length
* 59706 63841: contig of 4136 bp in length
* 63842 63855: gap of unknown length
* 63856 70264: contig of 6409 bp in length
* 70265 70278: gap of unknown length
* 70279 75486: contig of 5208 bp in length
*

```


AUTHORS Waterston,R.H.
TITLE The sequence of Caenorhabditis elegans clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 306131)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces * is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will be preserved.

*	1	28533:	contig of 28533 bp in length
*	28534	28548:	gap of unknown length
*	28549	63879:	contig of 35331 bp in length
*	63880	63894:	gap of unknown length
*	63895	306131:	contig of 242237 bp in length.

FEATURES
Location/Qualifiers
source 1..306131
/db_organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y54E10"

BASE COUNT	101407	a	53175	c	52480	g	99039	t	30	others
ORIGIN										

Query Match	15.4%;	Score	154;	DB	19;	Length	306131;
Best Local Similarity	78.7%;	Pred.	No. 4.30e-70;				
Matches	259;	Conservative	0;	Mismatches	63;	Indels	7; Gaps 6;

```

Db 302066 AAATCATTTTCTTTTTAACCTAAATAATTCAACACTTTTTAGAATAATATTACGGGACAC 302125
      ||||| | | | ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||
Cp 1674 AATCCTGTGCCCTTTAAAGTGGAATGCCAAAATCTTGTAAT-TATTACGGAAACAC 1616

Db 302126 AAAATTCGAGAATCGGCATTGCCAAAAACATATTTGACGCGCAAATAATCTCGCAGCGAAA 302185
      ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||
Cp 1615 AAAATTCAGAGAATCGGTATTACAGT-CATATTTGGCGCGCAAATAATCTCGTAGCTAGA 1557

Db 302186 ACTACAGTAATTCCTTTAAATGACTACTGTAGTCTCTGTGTCGATTACGGTTTTTGCCTA 302245
      ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||
Cp 1556 ACTACAGTAATCCCTTTAAATGACTACTGTAGCG-TTGTGACGATTTACGGGTAT-CAAA 1499

Db 302246 ATTTGAATATTCATCGGTCTAAAAAAAAAAAATTGAGCCCCGTAAATCGACACAAGCGCTACA 302305
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cp 1498 ATTGAAAAAANAATAATTATTTTCAAAA--TTGAGCCCCGTAAATCGACACAAGCGCTACA 1441

Db 302306 GTAGTCCTTTAAGTTATTACAGTAGTTTTCGCTTCGAAATATTTTGGCGGTCAAATATGT 302365
      ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | ||||| | | |||||
Cp 1440 GTAGTCATTTAAGAATTACTGTAGTTTTCGCTACGAGATATTTTGC-CGTCAAATATGT 1382

Db 302366 TGTGCAATACGCATTCTCAGGATTTTGTG 302394
      || | | | | | | | | | | | | | | | | | | | | | |
Cp 1381 TGCGCAGTACGCATTCTCAGAATTTTGTG 1353

RESULT 11
LOCUS AC006663 32209 bp DNA HTG 23-FEB-1999
DEFINITION Caenorhabditis elegans clone H24K24, WORKING DRAFT SEQUENCE, 1 unordered pieces.
ACCESSION AC006663
NID g4263275
VERSION AC006663.1 GI:4263275
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
          Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
          Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis
REFERENCE 1 (bases 1 to 32209)
AUTHORS Waterston,R.H.
```


* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

*      1      29929: contig of 29929 bp in length
*      29930 29943: gap of unknown length
*      29944 140702: contig of 110759 bp in length.
FEATURES             Location/Qualifiers
     source           1..140702
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /clone="Y61A9L"
BASE COUNT    44775 a   25754 c   24599 g   45560 t   14 others
ORIGIN
Query Match          14.6%; Score 146; DB 20; Length 140702;
Best Local Similarity 83.8%; Pred. No. 2.77e-65;
Matches 212; Conservative 0; Mismatches 36; Indels 5; Gaps 5;

Db 67485 ATTACGAAATCCGAGAATCGGTGTTACACAACATATTTGACGAGCAAAAATATCTCGTAG 67544
        ||| ||||| ||||| ||||| | | ||||| ||||| ||||| ||||| ||||| |||||
QY 1351 ATCACAATAATCTGAGAAATGCGTACTGCGCAACATATTTGACG-GCAAAAATATCTCGTAG 1409

Db 67545 CGAAAACTACAGTAACCCTTAAATGAATACGGTAGCGCTTGTCGATTTTCGGGGCTC 67604
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1410 CGAAAACTACAGTAATCTCTAAATGACTACTGTAGCGCTTGTCGATTTACGGG-CTC 1468

Db 67605 GATTTCTCTAAATTAATTTTTTTTTTTATTAGGAATCGAGCCCCGTAATCGACACAAAGCG 67664
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1469 AATTTTGGAAATAATTTTTTTT-T-CGAAATTTTGATAACCCGTAATCGTCACAA-CGC 1526

Db 67665 TACAGTAGTCATTTTAAGAAATTAAGTACTGTAGTTTTCGCTACGAGATATTTTTCGGCGTCATAT 67724
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1527 TACAGTAGTCATTTTAAGGATTACTGTAGTTCTAGCTACGAGATATTTTTCGGCGCAAAT 1586

Db 67725 ATGT-TGTAATAC 67736
        ||||| ||||| |||||
QY 1587 ATGACTGTAATAC 1599

RESULT 13
LOCUS       AC006906 317781 bp DNA HTG 26-FEB-1999
DEFINITION Caenorhabditis elegans clone Y75B7, WORKING DRAFT SEQUENCE, 13
            unordered pieces.
ACCESSION   AC006906
NID         g4309898
VERSION     AC006906.2 GI:4309898
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
            Caenorhabditis elegans.
ORGANISM    Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 317781)
AUTHORS    Waterston,R.H.
TITLE       The sequence of Caenorhabditis elegans clone
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 317781)
AUTHORS    Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (24-FEB-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT
On Mar 1, 1999 this sequence version replaced gi:4263442.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
```



```
/gene="ced-3"
/number=7
7013. .7652
/gene="ced-3"
/number=8

BASE COUNT      2429 a 1455 c 1271 g 2498 t
ORIGIN

Query Match      14.5%; Score 145; DB 21; Length 7653;
Best Local Similarity 83.3%; Pred. No. 1.10e-64;
Matches 244; Conservative 0; Mismatches 39; Indels 10; Gaps 8;

Db 1342 TATTAAGCAATCACAAAATTCGAGAAATCGGTACTGCGCAACATATTGACG-GCAAAAAT 1400
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1629 TATTACGGAACACAAAATTCAGAGAATCGGTATTACAGT-CATATTGGCGGCAAAAAT 1571

Db 1401 ATCTCGTAGCGAAACTACAGTAATCTTTAAATGACTGACTGCTAGCGCTTGTCGATTT 1460
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1570 ATCTCGTAGCTAGAACTACAGTAATCTTTAAATGACTGACTGCTAGCG-TTGTGACGATTT 1512

Db 1461 ACGGCT--CAATTTTGAATAAATTTTCTTTCGAATTTTGATAACCCGTAATCGT 1518
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1511 ACGGTTATCAAAATTCGAAAAAATAATTTATTTTCAAAATTTGAG--CCCGTAAATCGA 1454

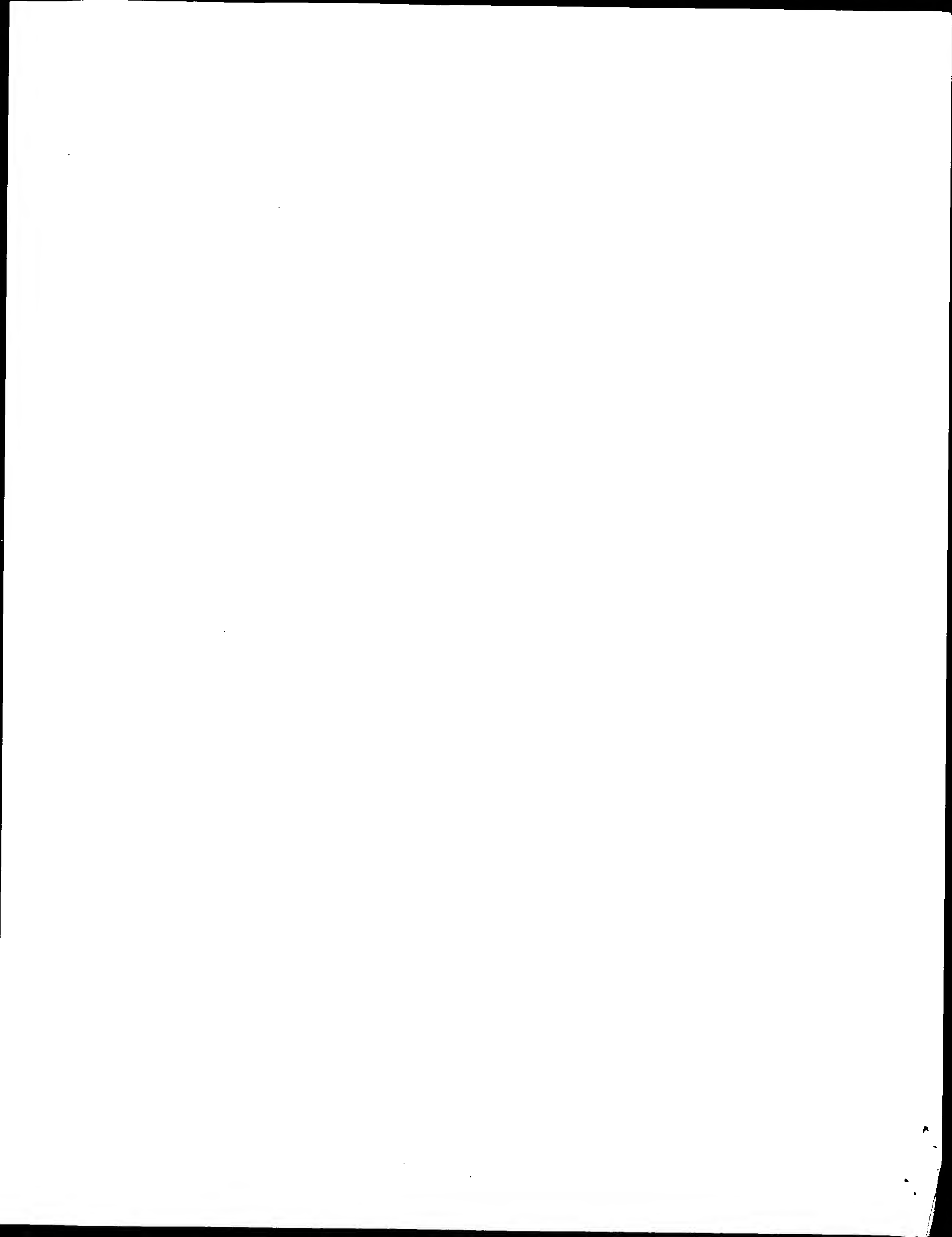
Db 1519 CACAA-CGCTACAGTAGTCATTTAAAGGATTACTGTAGTTCTAGCTACGAGATATTTGC 1577
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1453 CACAAGCGCTACAGTAGTCATTTAAAGAAATTACTGTAGTTTTCGCTACGAGATATTTGC 1394

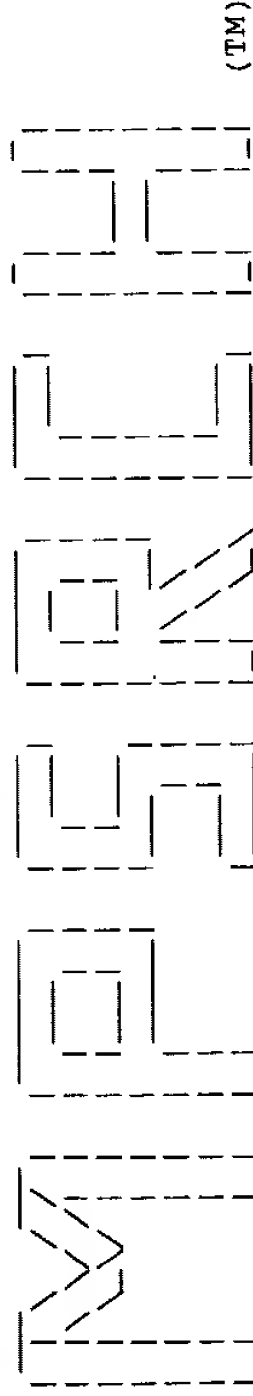
Db 1578 GCGCAAAATATGACT-GTAATACGCATTCTCTGAATTTTGTGTTTCCGTAATA 1629
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1393 -CGTCAATATGTTGCGCAGTACGCATTCTCAGAAATTTGTGATTCCCTTAATA 1342

RESULT 15
LOCUS      CEY49E10 125590 bp DNA INV 18-DEC-1998
DEFINITION Caenorhabditis elegans cosmid Y49E10, complete sequence.
ACCESSION  Z98866
NID         93217981
VERSION     Z98866.1 GI:3217981
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 125590)
AUTHORS     Barlow,K.
TITLE       Direct Submission
JOURNAL     Submitted (28-AUG-1997) Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rwenematode.wustl.edu
REFERENCE   2 (bases 1 to 125590)
AUTHORS     Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favellio,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifkin,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
TITLE       2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
JOURNAL     Nature 368 (6466), 32-38 (1994)
MEDLINE     94150718
COMMENT     On Jun 13, 1998 this sequence version replaced gi:2528901.
            Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=Y49E10

Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we arrange for a small
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
This sequence is the entire insert of clone Y49E10. The true left
end of clone Y111B2 is at 36231 in this sequence. The true right
end of clone Y75B8 is at 34585 in this sequence. The start of this
sequence (1..107) overlaps with the end of sequence AL033514.
Location/Qualifiers
1..125590
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="Y49E10"
complement(284..753)
/gene="Y49E10.6"
complement(join(284..412,472..753))
/gene="Y49E10.6"
/note="predicted using Genefinder; similar to Core histone
H2A/H2B/H3/H4; cDNA EST EMBL:D71193 comes from this gene;
cDNA EST Yk477a9.3 comes from this gene; cDNA EST
CEESG32RD comes from this gene; cDNA EST Yk201911.3 comes
from this gene; cDNA EST Yk333912.3 comes from this gene;
cDNA EST Yk376h11.3 comes from this gene; cDNA EST
EMBL:D73664 comes from this gene; cDNA EST EMBL:D72762
comes from this gene; cDNA EST CEESG32F comes from this
gene; cDNA EST Yk266a12.3 comes from this gene; cDNA EST
Yk266a12.5 comes from this gene; cDNA EST EMBL:D71904
comes from this gene; cDNA EST EMBL:D74642 comes from this
gene; cDNA EST EMBL:T02205 comes from this gene; cDNA EST
EMBL:T00023 comes from this gene; cDNA EST Yk458e6.3 comes
from this gene; cDNA EST Yk228d6.3 comes from this gene;
cDNA EST Yk481h9.3 comes from this gene; cDNA EST
Yk330a10.3 comes from this gene; cDNA EST Yk216b3.3 comes
from this gene; cDNA EST Yk477b11.3 comes from this gene;
cDNA EST Yk482e5.3 comes from this gene; cDNA EST
Yk482e5.5 comes from this gene; cDNA EST EMBL:D71687 comes
from this gene; cDNA EST Yk376b11.3 comes from this gene;
cDNA EST EMBL:T01965 comes from this gene; cDNA EST
Yk228h3.3 comes from this gene; cDNA EST Yk476b9.3 comes
from this gene; cDNA EST Yk319d5.3 comes from this gene;
cDNA EST EMBL:T01428 comes from this gene; cDNA EST
EMBL:T01640 comes from this gene; cDNA EST EMBL:T02166
comes from this gene; cDNA EST Yk484d6.3 comes from this
gene; cDNA EST Yk473h10.3 comes from this gene; cDNA EST
Yk215g5.3 comes from this gene; cDNA EST Yk402h8.3 comes
from this gene; cDNA EST Yk246c5.3 comes from this gene;
cDNA EST Yk246f4.3 comes from this gene"
/codon_start=1
/protein_id="CAB11546.1"
/db_xref="PID:el298479"
/db_xref="PID:g3979986"
/db_xref="GI:3979986"
/translation="MARTKQTARKSTGGKAPRQLATKAARKSAPTGGVKKPHRYRP
GTVALREIRRYQKSTELLIRKLPFORLVREIAQDFKTLRFQSAAGALQEAAYLV
GLFEDTNLCALHAKRVITMPKMDQLARRINGERA"
3057..5053
/gene="Y49E10.1"
join(3057..3182,3235..4008,4703..5053)
/gene="Y49E10.1"
/note="similar to ATPases associated with various cellular
activities (AAA); cDNA EST Yk411g8.5 comes from this gene;
cDNA EST Yk361e3.5 comes from this gene; cDNA EST
EMBL:Cl3357 comes from this gene; cDNA EST EMBL:D74251
comes from this gene; cDNA EST EMBL:D68737 comes from this
gene; cDNA EST Yk196b7.5 comes from this gene; cDNA EST
Yk440d10.5 comes from this gene; cDNA EST Yk453f1.5 comes
from this gene; cDNA EST Yk231f1.5 comes from this gene;
```


Search completed: Fri Aug 6 02:01:17 1999
Job time : 3680 secs.





Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Fri Aug 6 03:00:19 1999; MasPar time 219.29 Seconds
Tabular output not generated. 977.118 Million cell updates/sec

Title: >US-08-287-669-18
Description: (901-1900) from US08287669.seq (2 of 10)
Perfect Score: 1000
N.A. Sequence: 901 GTCGTCCTGGTATCCTCAA.....TGAAAAATTGTTTATGTT 1900
Comp: CAGCAGGAACCATAGGAGTT.....ACTTTTACACAAATACAA

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 8.878; Variance 6.553; scale 1.355

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	998	99.8	7653	24	Nematode Ced-3 gene.	0.00e+00
2	996	99.6	7653	9	ced-3 gene.	0.00e+00
3	992	99.2	7653	9	ced-3 (G2487A) gene.	0.00e+00
4	992	99.2	7653	9	ced-3 (C5940T) gene.	0.00e+00
5	992	99.2	7653	9	ced-3 (G6372A) gene.	0.00e+00
6	992	99.2	7653	9	ced-3 (G6536A) gene.	0.00e+00
7	992	99.2	7653	9	ced-3 (G6297A) gene.	0.00e+00
8	992	99.2	7653	9	ced-3 (C7020T) gene.	0.00e+00
9	992	99.2	7653	9	ced-3 (C6434T) gene.	0.00e+00

10	992	99.2	7653	9	Q64743	ced-3 (C6485T) gene.	0.00e+00
11	992	99.2	7653	9	Q64740	ced-3 (C6322T) gene.	0.00e+00
12	992	99.2	7653	9	Q54666	ced-3 gene.	0.00e+00
13	992	99.2	7653	9	Q64737	ced-3 (G5757A) gene.	0.00e+00
14	145	14.5	7653	24	T38196	Nematode Ced-3 gene.	1.27e-62
15	143	14.3	7653	9	Q64744	ced-3 (G6536A) gene.	1.64e-61
16	143	14.3	7653	9	Q64743	ced-3 (C6485T) gene.	1.64e-61
17	143	14.3	7653	9	Q64737	ced-3 (G5757A) gene.	1.64e-61
18	143	14.3	7653	9	Q64745	ced-3 (C7020T) gene.	1.64e-61
19	143	14.3	7653	9	Q64740	ced-3 (C6322T) gene.	1.64e-61
20	143	14.3	7653	9	Q64739	ced-3 (G6297A) gene.	1.64e-61
21	143	14.3	7653	9	Q64738	ced-3 (C5940T) gene.	1.64e-61
22	143	14.3	7653	9	Q64736	ced-3 (G2487A) gene.	1.64e-61
23	143	14.3	7653	9	Q64742	ced-3 (C6434T) gene.	1.64e-61
24	143	14.3	7653	9	Q54666	ced-3 gene.	1.64e-61
25	143	14.3	7653	9	Q64741	ced-3 (G6372A) gene.	1.64e-61
26	141	14.1	7653	9	Q54401	ced-3 gene.	2.11e-60
27	81	8.1	6560	9	Q54629	Genomic region contai	9.82e-28
28	76	7.6	6560	9	Q54629	Genomic region contai	4.29e-25
29	41	4.1	91	9	Q51746	Oligonucleotide probe	2.06e-07
30	40	4.0	204	1	N81164	Base substituted E.co	6.13e-07
31	36	3.6	91	9	Q51746	Oligonucleotide probe	4.45e-05
32	36	3.6	204	1	N81164	Base substituted E.co	4.45e-05
33	35	3.5	91	46	V44650	Mammalian DNA replica	1.27e-04
34	33	3.3	91	46	V44650	Mammalian DNA replica	1.02e-03
35	32	3.2	501	3	N50023	Sequence encoding new	2.84e-03
36	30	3.0	114	12	Q70469	Generic DNA sequence	2.14e-02
37	30	3.0	501	3	N50026	Sequence encoding new	2.14e-02
38	30	3.0	501	3	N50024	Sequence encoding new	2.14e-02
39	30	3.0	501	3	N50027	Sequence encoding new	2.14e-02
40	28	2.8	114	12	Q70467	Generic DNA sequence	1.54e-01
41	28	2.8	219	19	T20616	Human gene signature	1.54e-01
42	28	2.8	501	3	N50031	Sequence encoding new	1.54e-01
43	28	2.8	501	3	N50025	Sequence encoding new	1.54e-01
44	28	2.8	501	3	N50033	Sequence encoding new	1.54e-01
45	28	2.8	501	3	N50029	Sequence encoding new	1.54e-01

ALIGNMENTS

RESULT 1
ID T38196 standard; DNA; 7653 BP.
AC T38196;
DT 17-DEC-1996 (first entry)
DE Nematode Ced-3 gene.
KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;
KW cell death; apoptosis; neural degeneration; inflammation;
KW antiinflammatory; ds.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /tag= a
FT /label= Repeat-1
FT 1490..1614
FT /tag= b
FT /label= Repeat-2
FT 2167..2366
FT /tag= c
FT /codon_start= 2232..2234
FT allele 2310
FT /tag= d
FT /label= T(n1040)
FT /note= "causes L27F mutation"
FT intron 2367..2429
FT /tag= e
FT /label= Intron-1
FT 2430..2575
FT /tag= f
FT allele 2487
FT /tag= g
FT /label= A(n718)
FT /note= "causes G65R mutation"
FT intron 2576..2853


```

FT FT /*tag= h
FT FT /label= Intron-2
FT FT 2854..3107
FT FT /*tag= i
FT FT 3108..4302
FT FT /*tag= j
FT FT /label= Intron-3
FT FT 3126..3243
FT FT /*tag= k
FT FT /label= Repeat-1
FT FT 3329..3396
FT FT /*tag= l
FT FT /label= Repeat-1
FT FT 3487..3759
FT FT /*tag= m
FT FT /label= Repeat-2
FT FT 3782..4070
FT FT /*tag= n
FT FT /label= Repeat-2
FT FT 4303..4633
FT FT /*tag= o
FT FT 4634..5546
FT FT /*tag= p
FT FT /label= Intron-4
FT FT 4688..4719
FT FT /*tag= q
FT FT /label= Repeat-3
FT FT 5221..5330
FT FT /*tag= r
FT FT /label= Repeat-3
FT FT 5546..5760
FT FT /*tag= s
FT FT 5757
FT FT /*tag= t
FT FT /label= A(n2433)
FT FT /note= "causes G360S mutation"
FT FT 5761..5814
FT FT /*tag= u
FT FT /label= Intron-5
FT FT 5815..5942
FT FT /*tag= v
FT FT 5940
FT FT /*tag= w
FT FT /label= T(n1165)
FT FT /note= "creates premature stop codon at 403"
FT FT 5943..6297
FT FT /*tag= x
FT FT /label= Intron-6
FT FT 6062..6138
FT FT /*tag= y
FT FT /label= Repeat-4
FT FT 6298..6537
FT FT /*tag= z
FT FT 6322
FT FT /*tag= aa
FT FT /label= T(n1949)
FT FT /note= "creates premature stop codon at 412"
FT FT 6372
FT FT /*tag= ab
FT FT /label= A(n1286)
FT FT 6434
FT FT /*tag= ac
FT FT /label= T(n1129,n1164)
FT FT /note= "causes A449V mutation"
FT FT 6485
FT FT /*tag= ad
FT FT /label= T(n2430)
FT FT /note= "causes A466V mutation"
FT FT 6535
FT FT /*tag= ae
FT FT /label= A(n2426)
FT FT /note= "causes E483K mutation"
FT FT 6538..7012
FT FT intron

```

```

FT FT /*tag= af
FT FT /label= Intron-7
FT FT 6567..6625
FT FT /*tag= ag
FT FT /label= Repeat-5
FT FT 6905..6965
FT FT /*tag= ah
FT FT /label= Repeat-5
FT FT 7013..7653
FT FT /*tag= ai
FT FT /note= "in-frame stop codon at 7073-7075"
FT FT 7020
FT FT /*tag= aj
FT FT /label= T(n1163)
FT FT /note= "causes S486F mutation"
FT FT
PN WO9625946-A1.
PD 29-AUG-1996.
PF 23-FEB-1996; U02473.
PR 24-FEB-1995; US-394189.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WFL; 96-425082/42.
DR P-PSDB; R98754.
PT Ced-3 and human interleukin 1-beta convertase genes and proteins
PT useful to treat inflammation and diseases characterised by cell
PT death
PT
PS Claim 18; Fig 3; 139pp; English.
CC The Caenorhabditis elegans ced-3 gene (T38196) was cloned by
CC mapping RFLPs and chromosome walking, and genomic DNA cloned
CC in plasmid pJ107 was sequenced. EMS-induced alleles were also
CC sequenced. The gene codes for a cell death protein (R98754) that
CC is structurally similar to human interleukin-1 beta converting
CC enzyme (ICE) (R98755), suggesting that Ced-3 protein may be a
CC cysteine protease like ICE and that ICE may be a human equivalent
CC of the nematode cell death gene. The ced-3 gene can be used as a
CC probe or in the prodn. of Ced-3 protein and novel drugs for
CC enhancing or inhibiting the activity of ICE, ced-3 and related
CC genes for the treatment of inflammatory diseases and/or diseases
CC caused by cell death. Novel inhibitors of ced-3 activity include
CC portions of the ced-3 gene and its product.
SQ Sequence 7653 BP; 2429 A; 1456 C; 1270 G; 2498 T;

Query Match 99.8%; Score 998; DB 24; Length 7653;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 901 gtcgtccttggtatcctcaactgtcccggtttcttttcgggtacactcttcggtgatgc 960
QY |||||
901 GTCGTCCTGGTATCCTCAACTGTCCCGGTTTGTTCGGGTACACTCTCCGTGATGC 960

Db 961 cacctgtcctcgtctcaattatcggttttagaattggaactgtccagatgggtgactcata 1020
QY |||||
961 CACCTGTCTCCGTCCTCAATTATCGTTTAGAAATGTGAACGTGCCAGATGGGTGACTCATA 1020

Db 1021 ttgctgctgtacaaatccactttctttctcctcggcagtccttacgagcccatcaataac 1080
QY |||||
1021 TTGCTGCTGCTACAAATCCACTTCTTTCTCTCATCGGCAGTCTTACGAGCCCATCAAAAC 1080

Db 1081 ttttttttcgcgcgaaatttgcaataaaaccggcgaataaaactttctccaaattgttacgcaa 1140
QY |||||
1081 TTTTTCCTCCGCGAAATTTGCAATAAACCCGGCAAAACTTTCTCCAAATGTGTACGCAA 1140

Db 1141 tatatacaatccataagaatatcttctcaattgtttatgatttcttcgagcactttctct 1200
QY |||||
1141 TATATACAAATCCATAAGAATATCTTCTCAATGTTTATGATTCTTCTCGCAGCACTTTCTCT 1200

Db 1201 tcgtgtgctaacatcttattttataataatttcgcgtataaaattccgatttttgagtatta 1260
QY |||||
1201 TCGTGTGCTAACATCTTATTTTATAATATTCGCTAAATTCGATTTTTCGATTTTTCGATTTA 1260

Db 1261 atttatcgtaaaattatcataatagcacggaaaaactactaaaaaaggtaaaagctccttt 1320
QY |||||
1261 ATTTATCGTAAATATATCATATAATAGCACCGGAAAACTACTAAAAATGGTAAAGCTCCTTT 1320

```


Db	1321	taaatcggtcgacattatcgattattaaggaatacaaaaattctgagaatcggtactgcgc	1380	FT	mutation	/number= Exon_2
Qy	1321	TAAATCGGCTCGACATTATCGTATTAAGGAATCACAAAATCTGAGATCGGTACTGCGC	1380	FT	intron	/tag= g
Db	1381	aacatattgacggcaaaaatatctgtagcgaaaaactacagtaattctttaaatgactac	1440	FT	exon	/note= "G>A, fromm allele n718"
Qy	1381	AACATATTGACGGCAAAATATCTCGTAGCGAAACTACAGTAATCTTTTAAATGACTAC	1440	FT	intron	2576..2853
Db	1441	tgtagcgttggtgcgatttacgggctcaatttttgaataaatttttttttcgaattt	1500	FT	repeat_unit	/tag= h
Qy	1441	TGTAGCGTTGTTGTCGATTTACGGGCTCAATTTTGAATAAATTTTTTTTTCGAATTT	1500	FT	repeat_unit	/number= Intron_2
Db	1501	tgataaccgttaaatcgtaacacgctacagtagtcatttaaaaggattactgtattcta	1560	FT	intron	2854..3107
Qy	1501	TGATAACCCGTAAATCGTCACACGCTACAGTAGTCAATTTAAAGGATTACTGTAGTTCTA	1560	FT	repeat_unit	/tag= i
Db	1561	gctacgagataatttgcgcgcaaatatgactgtaatacgcattctctgaattttgtgtt	1620	FT	repeat_unit	/number= Exon_3
Qy	1561	GCTACGAGATATTTGCGCGCCAAATATGACTGTAATACGCATTCTCTGAATTTGTGTT	1620	FT	repeat_unit	3108..4302
Db	1621	tccgtaataatttcacaagattttggtcattccactttaaaggcgacaggattttatcca	1680	FT	repeat_unit	/tag= j
Qy	1621	ICCGTAATAATTTCAAGAATTTTGGCATTCACCTTTAAAGCGCACAGGATTTATTTCCA	1680	FT	repeat_unit	/number= Intron_3
Db	1681	atgggtctcgacgcaaaaaagttttagtagactttttaaattctccttgcatttttaattc	1740	FT	repeat_unit	3126..3243
Qy	1681	ATGGGTCTCGGCACGCAAAAAGTTTGTAGACTTTTAAATCTCCTTGCAATTTTAATTC	1740	FT	repeat_unit	/tag= k
Db	1741	aattactaaaaatttcgtgaatttttctgttaaaaatttttaaaatcagtttttcaatt	1800	FT	exon	/rpt_type= INVERTED
Qy	1741	AATTACTAAAATTTTCGTGAATTTTTCGTAAAATTTTAAATCAGTTTCTCTAATATT	1800	FT	exon	/note= "Inverted w.r.t. repeat at 3329-3396"
Db	1801	ttccaggctgacaaacagaaacaaaacacacacacacacacacacacacacacacacac	1860	FT	intron	/tag= l
Qy	1801	TTCCAGGCTGACAAACAGAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAAT	1860	FT	repeat_unit	/rpt_type= INVERTED
Db	1861	taaaaaataacgattttctcattgaaaattgtgttttatgtt	1900	FT	repeat_unit	/note= "Inverted w.r.t. repeat at 3782-4070"
Qy	1861	TAAAAATAACGATTTCTCATTGAAAATTTGTGTTTATGTT	1900	FT	exon	/tag= m
RESULT	2					
ID	Q54401	standard; DNA; 7653 BP.				
AC	Q54401;					
DT	01-JUL-1994	(first entry)				
DE	ced-3 gene.					
KW	C.elegans; cell death; gene; ced-3; inhibition; human; parasite;					
KW	interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;					
KW	inflammatory response; nematode; diagnosis; myocardial infarction;					
KW	stroke; degenerative disease; traumatic brain injury; hypoxia; pest;					
KW	pathogenic infection; hair loss; cancer; autoreactive antibody; ss.					
OS	Caenorhabditis elegans.					
PH	Key	Location/Qualifiers				
FT	repeat_unit	1356..1472				
FT		/tag= a				
FT		/rpt_type= INVERTED				
FT		/note= "Inverted w.r.t repeat at 1490-1614"				
FT	repeat_unit	1490..1614				
FT		/tag= b				
FT		/rpt_type= INVERTED				
FT		/note= "Inverted w.r.t. repeat at 1356-1472"				
FT	exon	2232..2366				
FT		/tag= c				
FT		/number= Exon_1				
FT	mutation	2310				
FT		/tag= d				
FT		/note= "C>T, from allele n1040"				
FT	intron	2367..2429				
FT		/tag= e				
FT		/number= Intron_1				
FT	exon	2430..2575				
FT		/tag= f				

mutation 6372
 /*tag= ac
 /note= "G>A, from allele n1286"
 6434
 /*tag= ad
 /note= "C>T, from alleles n1129 and n1164"
 6485
 /*tag= ae
 /note= "C>T, from allele n2430"
 6535
 /*tag= af
 /note= "G>A, from allele n2426"
 6538..7012
 /*tag= ag
 /number= Intron_7
 6567..6625
 /*tag= ah
 /rpt_type= INVERTED
 /note= "Inverted w.r.t. repeat at 6905-6965"
 6905..6965
 /*tag= ai
 /rpt_type= INVERTED
 /note= "Inverted w.r.t. repeat 6567-6625"
 7013..7075
 /*tag= aj
 /number= Exon_8
 7020
 /*tag= ak
 /note= "C>T, from allele n1163"
 1..5850
 /*tag= al
 /note= "Claim 2, inhibitor fragment"
 1..3020
 /*tag= am
 /note= "Claim 2, inhibitr fragment"
 WO9325694-A.
 23-DEC-1993.
 14-JUN-1993; U05705.
 12-JUN-1992; US-897788.
 20-NOV-1992; US-984182.
 (MASI) MASSACHUSETTS INST TECHNOLOGY.
 Horvitz HR, Shahan S, Yuan J;
 WPI; 94-007551/01.
 P-PSDB; R45262.
 Agents which affect activity of cell death genes - used to
 develop drugs for treating conditions characterised by cell death
 or proliferation
 Disclosure; Fig 3; 132pp; English.
 This sequence encodes the C.elegans cell death gene, ced-3. Fragments
 of the amino terminal of the protein encoded by this sequence act as
 inhibitors of ced-3. This gene has considerable similarity to human
 interleukin-1beta convertase (ICE), which converts pro-interleukin-
 1beta to the active cytokine and is involved in inflammatory response
 in humans. The similarity between the two sequences suggests that
 inhibitors of ced-3 may also act as inhibitors of ICE. Human ICE and
 nematode Ced-3 proteins have an overall amino acid similarity of 28%.
 The ced-3 inhibitors may be used for identifying agents which affect
 the activity of a gene belonging to the to the ced-3/ICE family of
 genes and for diagnosis of diseases characterised by cell death. They
 can also be used to develop drugs for treating conditions characterised
 by cell deaths such as myocardial infarction, stroke, degenerative
 disease, traumatic brain injury, hypoxia, pathogenic infection, or
 hair loss, or drugs for reducing the proliferative capacity or size
 of a population of cells such as cancerous cells, cells which produce
 autoreactive antibodies, infected cells, hair follicle cells or cells
 which are critical to the life of a parasite, pest or recombinant
 organism. They may also be used in the diagnosis of inflammatory
 disease.
 Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.6%; Score 996; DB 9; Length 7653;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db	901	gtcgtccttggtatctcctcaacttgcctccggtttgttttcggtacactcttcggtgatgc	960
QY	901	gtcgtccttggtatcctcaacttgcctccggtttgttttcggtacactcttcggtgatgc	960
Db	961	cacctgtcctcgtctcaattatcgttttagaaatgtgaactgtccagatgggtgactcata	1020
QY	961	cacctgtcctcgtctcaattatcgttttagaaatgtgaactgtccagatgggtgactcata	1020
Db	1021	ttgctgtgctacaaatccactttctttctcctcgtcggcagctcttaccagcccatcataaac	1080
QY	1021	ttgctgtgctacaaatccactttctttctcctcgtcggcagctcttaccagcccatcataaac	1080
Db	1081	tttttttccgcgaaatttgcaataaaacccggcccaaaaactttctccaaattgttaccgcaa	1140
QY	1081	tttttttccgcgaaatttgcaataaaacccggcccaaaaactttctccaaattgttaccgcaa	1140
Db	1141	tatatcaaatccataagaatatcttctcctcaatgtttatgatttcttcgagcactttctct	1200
QY	1141	tatatcaaatccataagaatatcttctcctcaatgtttatgatttcttcgagcactttctct	1200
Db	1201	tcgtgtgctaacaatcttttttataataatttcgcgtaaaaattcctttaaattgagtatta	1260
QY	1201	tcgtgtgctaacaatcttttttataataatttcgcgtaaaaattcctttaaattgagtatta	1260
Db	1261	atttatcgtaaaaatcataatagcacgcaaaaactactaaaaatggtaaaagctccttt	1320
QY	1261	atttatcgtaaaaatcataatagcacgcaaaaactactaaaaatggtaaaagctccttt	1320
Db	1321	taaatcggctcgacattatcgtatttaagggaatcacaaaattctggagaatgcgtactgcgc	1380
QY	1321	taaatcggctcgacattatcgtatttaagggaatcacaaaattctggagaatgcgtactgcgc	1380
Db	1381	aacataatttgacgtcaaaaatatctcgtagggaataactacagtaattctttaaatactac	1440
QY	1381	aacataatttgacgtcaaaaatatctcgtagggaataactacagtaattctttaaatactac	1440
Db	1441	tgtagcgttggtgctgatttacggggtcctcaatttttgaaaaataatttttttcgaattt	1500
QY	1441	tgtagcgttggtgctgatttacggggtcctcaatttttgaaaaataatttttttcgaattt	1500
Db	1501	tgataaccggtaaatcgtcacaaagctcacagtagtgcattttaaaggattactgtagtcta	1560
QY	1501	tgataaccggtaaatcgtcacaaagctcacagtagtgcattttaaaggattactgtagtcta	1560
Db	1561	gctacgagatattttgcgcgcaaatatgactgtataatcgcgacttctctgaattttgtgtt	1620
QY	1561	gctacgagatattttgcgcgcaaatatgactgtataatcgcgacttctctgaattttgtgtt	1620
Db	1621	tcgtaataatttcacaaagattttggcattccacttttaaaattctccttgcatatttccca	1680
QY	1621	tcgtaataatttcacaaagattttggcattccacttttaaaattctccttgcatatttccca	1680
Db	1681	atgggtctcggcagcaaaaagtttgatagactttttaaattctccttgcatatttccca	1740
QY	1681	atgggtctcggcagcaaaaagtttgatagactttttaaattctccttgcatatttccca	1740
Db	1741	aattactaaaaattttcgtgaatttttctgttaaaatttttaaattcagtttttctaatatt	1800
QY	1741	aattactaaaaattttcgtgaatttttctgttaaaatttttaaattcagtttttctaatatt	1800
Db	1801	ttccagggtgacaaacagaaaaac	1860
QY	1801	ttccagggtgacaaacagaaaaac	1860
Db	1861	taaaaaataacgattttctcattgaaaaattgtgttttatgtt	1900
QY	1861	taaaaaataacgattttctcattgaaaaattgtgttttatgtt	1900

RESULT 3
 ID Q64736 standard; DNA; 7653 BP.
 AC Q64736;

23-JUN-1994 (first entry)
ced-3 (G2487A) gene.
C. elegans; ced-3; mutant; transcriptional regulation;
embryogenesis; cell death; hydrophilic; transmembrane; region;
hydrophobic; mutation; amino acid; substitution; RNA splicing;
protein synthesis; null phenotype; calcium-binding domain; ss.
Caenorhabditis elegans.
Key Location/Qualifiers
repeat_unit 1356..1472
/*tag= a
/rpt_type= INVERTED
/note= "Inverted w.r.t repeat at 1490-1614"
repeat_unit 1490..1614
/*tag= b
/rpt_type= INVERTED
/note= "Inverted w.r.t. repeat at 1356-1472"
exon 2232..2366
/*tag= c
/number= Exon_1
2367..2429
/*tag= d
/number= Intron_1
2450..2575
/*tag= e
/number= Exon_2
2487
/*tag= f
/note= "G>A, from allele n718"
intron 2576..2853
/*tag= g
/number= Intron_2
2854..3107
/*tag= h
/number= Exon_3
3108..4302
/*tag= i
/number= Intron_3
3126..3243
/*tag= j
/rpt_type= INVERTED
/note= "Inverted w.r.t. repeat at 3329-3396"
repeat_unit 3329..3396
/*tag= k
/rpt_type= INVERTED
/note= "Inverted w.r.t. repeat at 3126-3243"
repeat_unit 3487..3759
/*tag= l
/rpt_type= INVERTED
/note= "Inverted w.r.t. repeat at 3782-4070"
repeat_unit 3782..4070
/*tag= m
/rpt_type= INVERTED
/note= "Inverted w.r.t. repeat 3487-3759"
exon 4303..4634
/*tag= n
/number= Exon_4
4635..5546
/*tag= o
/number= Intron_4
4688..4719
/*tag= p
/rpt_type= INVERTED
/note= "Repeat 3"
repeat_unit 5221..5330
/*tag= q
/rpt_type= INVERTED
/note= "Repeat 3"
exon 5547..5760
/*tag= r
/number= Exon_5
5761..5814
/*tag= s
/number= Intron_5

exon 5815..5942
/*tag= t
/number= Exon_6
5943..6297
/*tag= u
/number= Intron_6
6062..6138
/*tag= v
/rpt_type= INVERTED
6298..6537
/*tag= w
/number= Exon_7
6538..7012
/*tag= x
/number= Intron_7
6567..6625
/*tag= y
/rpt_type= INVERTED
/note= "Inverted w.r.t. repeat at 6905-6965"
repeat_unit 6905..6965
/*tag= z
/rpt_type= INVERTED
/note= "Inverted w.r.t. repeat 6567-6625"
exon 7012..7075
/*tag= aa
/number= Exon_8
WO9325685-A.
23-DEC-1993.
14-JUN-1993; U05701.
12-JUN-1992; US-897788.
20-NOV-1992; US-979638.
(MASI) MASSACHUSETTS INST TECHNOLOGY.
Horvitz HR, Shaham S, Yuan J;
WPI; 94-007542/01.
P-PSDB; R53280.
Isolated C elegans cell death genes ced-3 and ced-4 - used to
develop agents to increase or prevent cell death in organisms
Claim 14; Fig 4; 127pp; English.
The sequences given in Q64735-45 represent mutations of the C. elegans
ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
was most abundant in embryos, but was also detected in larvae and young
adults, suggesting that ced-3 is expressed not only in cells undergoing
cell death. The four largest introns as well as sequences 5' of the
start codon contain repetitive elements, some of which have been
characterised in non-coding regions of other C. elegans genes, such
as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
length. Ced-3 is highly hydrophilic with no significant hydrophobic
region that might be a transmembrane region. One region of Ced-3 is
very rich in serine. It is thought that this region is involved in
protein-protein interactions, similar to acid blobs in transcription
factors. Of the mutations which occur within the ced-3 gene, eight of
the mutations are missense mutations, two are nonsense mutations and
two are putative splicing mutations. These mutations establish the
null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
function is not essential for viability. The ced-3 and ced-4 gene
products may be used to develop agents for treating conditions
characterised by cell deaths, such as myocardial infarction, stroke,
degenerative disease, traumatic brain injury, hypoxia, pathogenic
infection, aging or hair loss.
SQ Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;
Query Match 99.2%; Score 992; DB 9; Length 7653;
Best Local Similarity 99.6%; Pred. No. 0.00e+00;
Matches 996; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 901 gtcgtccttggtatcctcaactgtccccggttttgcgttacacttccgtgatgc 960
|||||
QY 901 GTCGTCCTTGGTATCCTCAACTGTCCCGGTTTGTTCGGTACACTCTCCGTGATGC 960
|||||
Db 961 cacctgtctccgtctcattatcggttagaataatgtaactgtccagatgggtgactcata 1020
|||||
QY 961 CACCTGTCTCCGTCCTCAATTATCGTTTAGAAATGTGAAGTGTCCAGATGGGTGACTCATA 1020
|||||

QY	1141	TATATACAATCCATAAAGAAATATCTTCTCAATGTTTATGATTTCTTCGCAGCACTTTCTCT	1200
Db	1201	tcgtgtgctaacaatcttattttttataataattccgcgtaaaaatccogatttttgagattta	1260
QY	1201	TCGTGTGCTAACATCTTATTTTATAATATATTTCCGCTAAAATCCGATTTTGTGAGTATTA	1260
Db	1261	atttatcgtaaaattatgataatagcaccgaaactactaaaaatggtaaaaagctccttt	1320
QY	1261	ATTATCGTAAAAATTATCATAATAGCACCGAAAACTACTAAAAATGGTAAAAGCTCCTTT	1320
Db	1321	taaatcggtcgacattatcgattataaggaatcacaaaaattctgagaatgcgtactgcgc	1380
QY	1321	TAAATCGGCTCGACATTATCGTATTAAAGGAATCACAAAATTC TGAGAATGCGTACTGCGC	1380
Db	1381	aacataattgacggcaaaaatatctcgtagcgaaaaactacagtaattcttttaaatgactac	1440
QY	1381	AACATATTTGACGGCAAAAATATCTCGTAGCGAAAACTACAGTAATCTTTTAAATGACTAC	1440
Db	1441	tgtagcgcttgtgcgatttacgggctcaatttttgaaaaataatttttttttcgaaatt	1500
QY	1441	TGTAGCGCTTGTGTCGATTTACGGGCTCAATTTTGAAAAATAATTTTTCGAAATTT	1500
Db	1501	tgataacccgtaaaatcgtcacaaacgctacagtagtcattttaaggattactgtagttcta	1560
QY	1501	TGATAACCCGTAATAATCGTCACACGCTACAGTAGTCATTTAAAGGATTACTGTAGTTCTA	1560
Db	1561	gctaogacataattttgcgcgcccataatgactgtaatacgcattctctgaaatttttgtgtt	1620
QY	1561	GCTACGAGATATTTTGC GCGGCCAAAATATGACTGTAAATACGCATTTCTCTGAATTTTGTGTT	1620
Db	1621	tccgtaataatttcacagagattttggcattccactttaaggcgcaaggattttattcca	1680
QY	1621	TCCGTAATAAATTTACAAAGATTTTGGCATTTCCACTTTAAAGCGCACAGGATTTATTCCA	1680
Db	1681	atgggtctgggcacgnaaaagttttgatagacttttaaaattctccttgcatttttaattc	1740
QY	1681	ATGGGTCTCGGCACGCAAAAAGTTTGATAGACTTTTAAATTCCTCCITGCAATTTTAAATTC	1740
Db	1741	aattactaaaaatttcgtgaatttttctgtttaaaaaattttaaaatcagtttttctaataatt	1800
QY	1741	AATTACTAAAAATTTTCGTGAATTTTCTGTTAAAAATTTTAAAAATCAGTTTTCCTAATATT	1800
Db	1801	ttccaggtgacaaacagaaaaac	1860
QY	1801	TTCCAGGCTGACAAACAGAAAAACAAAAACACAAACACATTTTAAAAATCAGTTTTCAAAT	1860
Db	1861	taaaaaataacgatttctcattgaaaaattgtgttttatgtt	1900
QY	1861	TAAAAATAACGATTTCTCATTTGAAAAATTTGTGTTTATGTT	1900

	Query Match	99.2%;	Score 992;	DB 9;	Length 7653;
	Best Local Similarity	99.6%;	Pred. No. 0.00e+00;		
	Matches 996;	Conservative	0;	Mismatches 4;	Indels 0;
					Gaps 0;
Db	901	gtcgtccttggtatcctcaactgtccccgggttttgttttcggtacactcttccggtgatgc	960		
QY	901	gtcgtccttggtatcctcaactgtccccgggttttgttttcggtacactcttccggtgatgc	960		
Db	961	cacctgtctccgtctcctaattatcgtttagaatagtgaactgtccagatgggtgactcata	1020		
QY	961	cacctgtctccgtctcctaattatcgtttagaatagtgaactgtccagatgggtgactcata	1020		
Db	1021	ttgctgtcgtctacaatccactttctttctctcatcggcagtccttacgagcccatcataaac	1080		
QY	1021	ttgctgtcgtctacaatccactttctttctctcatcggcagtccttacgagcccatcataaac	1080		
Db	1081	ttttttttccgcgaaatttgcaataaaccggcccaaaaaactttctccaaattgtttacgcaa	1140		
QY	1081	ttttttttccgcgaaatttgcaataaaccggcccaaaaaactttctccaaattgtttacgcaa	1140		
Db	1141	tatatcaaatcccataagaatatcttctcaatgtttatgatttcttcgcagcacatttctct	1200		

```

RESULT      5
ID   Q64741 standard; DNA; 7653 BP.
AC   Q64741;
DT   23-JUN-1994 (first entry)
DE   ced-3 (G6372A) gene.
KW   C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW   embryogenesis; cell death; hydrophilic; transmembrane; region;
KW   hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW   protein synthesis; null phenotype; calcium-binding domain; ss.
OS   Caenorhabditis elegans.
FH   Key Location/Qualifiers
FT   repeat_unit 1356..1472
FT               /*tag= a
FT               /rpt_type= INVERTED
FT               /note= "Inverted w.r.t repeat at 1490-1614"
FT   repeat_unit 1490..1614
FT               /*tag= b
FT               /rpt_type= INVERTED
FT               /note= "inverted w.r.t. repeat at 1356-1472"
FT   exon 2232..2366
FT               /*tag= c
FT

```


PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PK 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
PB WPI; 94-007542/01.
PK P-PSDB; R53287.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127pp; English.
CC The sequences given in Q64735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

Query Match 99.2%; Score 992; DB 9; Length 7653;
Best Local Similarity 99.6%; Pred. No. 0.00e+00;
Matches 996; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Db 901 gtcgtccttggtatcctcaactgtcccggtttgttttcggtacactcttcggtgatgc 960
|||
QY 901 GTCGTCTCTGGTATCCTCAACTGTCCGGTTTGTTCGGTACACTCTTCCGTGATGC 960
|||
Db 961 cactgtctcgtctcaattatcgttttagaataatgtgaactgtccagatgggtgactcata 1020
|||
QY 961 CACCTGTCTCGTCTCAATTATCGTTAGAAATGTGAACGTCCAGATGGGTGACTCATA 1020
|||
Db 1021 ttgctgtctgtacataccactttcttctctcatcggcagcttctacgagccatcataaac 1080
|||
QY 1021 TTGCTGTCTGTACAAATCCACTTTCTTTCTCATCGGCAGCTTTACGAGCCCATCAATAAC 1080
|||
Db 1081 ttttttttcgcgcgaatttgcaataaacacggcccaaaaactttctccaaaattgttaagcaa 1140
|||
QY 1081 TTTTTCCTCCGCGAAATTGCAATAAACCGGCCCAAACTTCTCCAAATTGTTACGCAA 1140
|||
Db 1141 tatatacaatccataagaatatcttctcaatgtttatgtatttcttcgcagcacttctct 1200
|||
QY 1141 TATATACAATCCATAAGAATATCTCTCAATGTTTATGATTTCTTCGACGCACTTCTCT 1200
|||
Db 1201 tcgtgtgtacacatcttattttataataatttcggtctaaaattccgatttttgagtatta 1260
|||
QY 1201 TCGTGTGCTAACATCTTATTTTATAATATTTCGCGTAAATTCGATTTTGTGATTA 1260
|||
Db 1261 attatgtataaattatgataatagcaccgaaactactataaaaatggtaaaagctccttt 1320
|||
QY 1261 ATTTATCGTAAATATTATCATAATAGCACCAGAACTACTATAAAATGGTAAAGCTCCTTT 1320
|||
Db 1321 taaatcggtcgcacattatcgtatttaagggaatcacaaaattctgagaatgggtactgcgc 1380
|||
QY 1321 TAAATCGGCTCGACATTATCGTATTAAAGGAATCACAAAATCTGAGAAATGCGTACTGCGC 1380
|||
Db 1381 aacatatttgacggcgaataatctcgtagcgaaactacagtaattctttaaataactactac 1440
|||
QY 1381 AACATATTGACGGCAAAATATCTCGTAGCGAAACTACAGTAATCTTTAAATGACTAC 1440
|||

Db 1441 ttagcgcgttggtgctgatttacggggtcgaatttttgaaaaataatttttttttcgaattt 1500
|||
QY 1441 TGTAGCGCTTGTGTGCTGATTACGGGCTCAATTTTGAATAATAATTTTTCGAATTT 1500
|||
Db 1501 tgataaccccgtaaatcgtcaaacgcgtacagtagtgcatttaaaaggattactgtattcta 1560
|||
QY 1501 TGATAACCCGTAATCGTCACAACGCTACAGTAGTCATTTAAAGGATTACTGTAGTTCTA 1560
|||
Db 1561 gctacgacataattttgctgcgcgaataatgactgtaataacgcatctctcgaattttgtgtt 1620
|||
QY 1561 GCTACGAGATATTTTGGCGGCCAAATATGACTGTAATACGCATCTCTGAAATTTTGTGTT 1620
|||
Db 1621 tcgtaataatttcacaagatttttggtcattccactttaaaggcgacaggtattattcca 1680
|||
QY 1621 TCCGTAATAATTTCAAGAATTTTGGCATTCACACTTTAAAGGCGCACAGGATTTATTTCCA 1680
|||
Db 1681 atgggtctgggcacggaataagtttgatagacttttaaaattctccttgcatcttttaattc 1740
|||
QY 1681 ATGGGTCTCGGCACGCAAAAAGTTTGTAGACTTTTAAATTTCTCTTCGATTTTAAATTC 1740
|||
Db 1741 aattactaaaattttcgtgaatttttctgttaaaatttttaaaatcagtttttctaatatt 1800
|||
QY 1741 AATTACTAAAATTTTCGTGAATTTTCTGTAAATTTTAAATCAGTTTCTTAATATT 1800
|||
Db 1801 ttccaggctgacaaacagaaacacaaacacacacacacacacacacacacacacacacacac 1860
|||
QY 1801 TTCCAGGCTGACAAACAGAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAT 1860
|||
Db 1861 taaaaataacgatttctcattgaaattgtgttttatgtt 1900
|||
QY 1861 TAAAAATAACGATTTCTCATTTGAAATTTGTGTTTATGTT 1900
|||

RESULT 7
ID Q64739 standard; DNA; 7653 BP.
AC Q64739;
DT 23-JUN-1994 (first entry)
DE ced-3 (G6297A) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472 /*tag= a
FT /*tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t repeat at 1490-1614"
FT repeat_unit 1490..1614 /*tag= b
FT /*tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366 /*tag= c
FT /*tag= c
FT /number= Exon_1
FT intron 2367..2429 /*tag= d
FT /*tag= d
FT /number= Intron_1
FT exon 2450..2575 /*tag= e
FT /*tag= e
FT /number= Exon_2
FT intron 2576..2853 /*tag= f
FT /*tag= f
FT /number= Intron_2
FT exon 2854..3107 /*tag= g
FT /*tag= g
FT /number= Exon_3
FT intron 3108..4302 /*tag= h
FT /*tag= h
FT /number= Intron_3
FT repeat_unit 3126..3243 /*tag= i
FT /*tag= i


```
QY 1561 GCTACGAGATATTTTCGGCGCCAAATATGACTGTAAATACGCATTCTCTGAATTTTGTTGTT 1620
Db 1621 tccgtaataatttcacaagattttggcattccacttttaaaagcgacaggaattattcca 1680
QY 1621 TCCGTAATAATTTCAAGATTTTGGCATTCCACTTTAAAGGGCGCAGGATTTATTCCA 1680
Db 1681 atgggtctcgccacgcaaaaagtttgatagacttttaaaattctcttgcatTTTtaattc 1740
QY 1681 ATGGGTCTCGGCACGCAAAAAGTTTGATAGACTTTTAAATTTCTCTGTCATTTTAAATTC 1740
Db 1741 aattactaaaatttcgtgaatttttctgttaaaatttttaaaatcagtttttctaatt 1800
QY 1741 AATTACTAAAATTTTCGTGAATTTTCTCTGTTAAATTTTAAATCAGTTTCTTAATAT 1800
Db 1801 ttccaggctgacaaacagaaacaaacaaacaaacaaacatttttaaaatcagtttttcaaat 1860
QY 1801 TTCCAGGCTGACAAACAGAAACAAACAAACAAACAAACATTTTAAATCAGTTTTCATAAT 1860
Db 1861 taaaaataacgatttctcattgaaattgtgtttttatgtt 1900
QY 1861 TAAAAATAACGATTTCTCATTGAAATTTGTGTTTATGTT 1900
```

RESULT 8

```
ID Q64745 standard; DNA; 7653 BP.
AC Q64745;
DT 23-JUN-1994 (first entry)
DE ced-3 (C7020T) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472 /*tag= a
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat at 1490-1614"
FT repeat_unit 1490..1614 /*tag= b
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366 /*tag= c
FT /*number= Exon_1
FT intron 2367..2429 /*tag= d
FT /*number= Intron_1
FT exon 2450..2575 /*tag= e
FT /*number= Exon_2
FT intron 2576..2853 /*tag= f
FT /*number= Intron_2
FT exon 2854..3107 /*tag= g
FT /*number= Exon_3
FT intron 3108..4302 /*tag= h
FT /*number= Intron_3
FT repeat_unit 3126..3243 /*tag= i
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit 3329..3396 /*tag= j
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat at 3126-3243"
FT repeat_unit 3487..3759 /*tag= k
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat at 3782-4070"
```

```
FT repeat_unit 3782..4070 /*tag= l
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat 3487-3759"
FT exon 4303..4634 /*tag= m
FT /*number= Exon_4
FT intron 4635..5546 /*tag= n
FT /*number= Intron_4
FT repeat_unit 4688..4719 /*tag= o
FT /*rpt_type= INVERTED
FT /*note= "Repeat 3"
FT repeat_unit 5221..5330 /*tag= p
FT /*rpt_type= INVERTED
FT /*note= "Repeat 3"
FT exon 5547..5760 /*tag= q
FT /*number= Exon_5
FT intron 5761..5814 /*tag= r
FT /*number= Intron_5
FT exon 5815..5942 /*tag= s
FT /*number= Exon_6
FT intron 5943..6297 /*tag= t
FT /*number= Intron_6
FT repeat_region 6062..6138 /*tag= u
FT /*rpt_type= INVERTED
FT exon 6298..6537 /*tag= v
FT /*number= Exon_7
FT intron 6538..7012 /*tag= w
FT /*number= Intron_7
FT repeat_unit 6567..6625 /*tag= x
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965 /*tag= y
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7012..7075 /*tag= z
FT /*number= Exon_8
FT mutation 7020 /*tag= w
FT /*note= "C>T, from allele n1163"
PN W09325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shahan S, Yuan J;
DR WPI; 94-007542/01.
DR P-PSDB; R53288.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127pp; English.
CC The sequences given in Q64735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The ced-3 protein is 503 amino acids in
```


length. Ced-3 is highly hydrophilic with no significant hydrophobic region that might be a transmembrane region. One region of Ced-3 is very rich in serine. It is thought that this region is involved in protein-protein interactions, similar to acid blobs in transcription factors. Of the mutations which occur within the ced-3 gene, eight of the mutations are missense mutations, two are nonsense mutations and two are putative splicing mutations. These mutations establish the null phenotype of the ced-3 gene, confirming that ced-3, like ced-4, function is not essential for viability. The ced-3 and ced-4 gene products may be used to develop agents for treating conditions characterised by cell deaths, such as myocardial infarction, stroke, degenerative disease, traumatic brain injury, hypoxia, pathogenic infection, aging or hair loss.

Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.2%; Score 992; DB 9; Length 7653;
Best Local Similarity 99.6%; Pred. No. 0.00e+00;
Matches 996; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 901 gtcgtccttggtatcctcaactgtcccggtttgttttgcgtacactctccgtgatgc 960
|||||
Qy 901 gtcgtccttggtatcctcaactgtcccggtttgttttgcgtacactctccgtgatgc 960
|||||
Db 961 caccgtctccgtctcaattatcgcttagaataatgtgaactgtccagatgggtgactcata 1020
|||||
Qy 961 caccgtctccgtctcaattatcgcttagaataatgtgaactgtccagatgggtgactcata 1020
|||||
Db 1021 ttgctgtcgtacaaaccactttctttctcatcggcagctctacagcccatcataaac 1080
|||||
Qy 1021 ttgctgtcgtacaaaccactttctttctcatcggcagctctacagcccatcataaac 1080
|||||
Db 1081 tttttttccgcgaaatttgcaataaaaccggcccaaaactttctccaaattgtttacgcaa 1140
|||||
Qy 1081 tttttttccgcgaaatttgcaataaaaccggcccaaaactttctccaaattgtttacgcaa 1140
|||||
Db 1141 tatatacaatccataagaatatctctcaatgtttatgatattcttcgcagcaactttctct 1200
|||||
Qy 1141 tatatacaatccataagaatatctctcaatgtttatgatattcttcgcagcaactttctct 1200
|||||
Db 1201 tcgtgtcgtacacatttttttataataatttcgcgtataaaattccgatttttgagtatta 1260
|||||
Qy 1201 tcgtgtcgtacacatttttttataataatttcgcgtataaaattccgatttttgagtatta 1260
|||||
Db 1261 atttatcgtaaaattatgataatagcagcgaataactactataaaatggtataagctccttt 1320
|||||
Qy 1261 atttatcgtaaaattatgataatagcagcgaataactactataaaatggtataagctccttt 1320
|||||
Db 1321 taaatcggctgcacattatcgtatttaaggaatacacaataattctgagaatgcgtactgcgc 1380
|||||
Qy 1321 taaatcggctgcacattatcgtatttaaggaatacacaataattctgagaatgcgtactgcgc 1380
|||||
Db 1381 aacatatttgacggcaaaatatctcgtagcgaaataactacagtaattctttaaataactac 1440
|||||
Qy 1381 aacatatttgacggcaaaatatctcgtagcgaaataactacagtaattctttaaataactac 1440
|||||
Db 1441 tgtagcgttgcgtgatttacgggctcaatttttgaaaaataatttttttttcgaattt 1500
|||||
Qy 1441 tgtagcgttgcgtgatttacgggctcaatttttgaaaaataatttttttttcgaattt 1500
|||||
Db 1501 tgataacccgtaaatcgtaacacgctacagtagtattttaaaaggattactgtagtcta 1560
|||||
Qy 1501 tgataacccgtaaatcgtaacacgctacagtagtattttaaaaggattactgtagtcta 1560
|||||
Db 1561 gctacgacataattttgcgcgcaaaatatgactgtaatacgcattctctgaattttgtgtt 1620
|||||
Qy 1561 gctacgacataattttgcgcgcaaaatatgactgtaatacgcattctctgaattttgtgtt 1620
|||||
Db 1621 tccgtaataatttcacaagattttggcattccacttttaaaggcgacaggaattttattcca 1680
|||||
Qy 1621 tccgtaataatttcacaagattttggcattccacttttaaaggcgacaggaattttattcca 1680
|||||
Db 1681 atgggtctggcagcaaaaagtgtgatagacttttaaaattctccttgcatttttaattc 1740
|||||

Qy 1681 ATGGGTCTCGGCACGCAAAAGTTTGATAGACTTTTAAATTCCTTGCATTTTAAATTC 1740
Db 1741 aattactaaaaatttctgtaatttttctgttaaaatttttaaaatcagtttttctaattt 1800
|||||
Qy 1741 AATTACTAAATTTTCGTGAATTTTCTGTAAATTTTAAATTCAGTTTCTTAATATT 1800
|||||
Db 1801 ttccaggctgacaaacagaaacacaaacacaaacacatttttaaaatcagtttttcaaat 1860
|||||
Qy 1801 TTCCAGGCTGACAAACAGAAACAAACAAACAAACAAATTTTAAATTCAGTTTCAAT 1860
|||||
Db 1861 taaaaataacgattttctcattgaaattgtgttttatgtt 1900
|||||
Qy 1861 TAAAAATAACGATTTCTCATTTGAAATTTGTGTTTATGTT 1900
|||||

RESULT 9
ID Q64742 standard; DNA; 7653 BP.
AC Q64742;
DT 23-JUN-1994 (first entry)
DE ced-3 (C6434T) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472 /*tag= a
FT /*tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t repeat at 1490-1614"
FT 1490..1614
FT /*tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT 2232..2366
FT /*tag= c
FT /number= Exon_1
FT 2367..2429
FT /*tag= d
FT /number= Intron_1
FT 2450..2575
FT /*tag= e
FT /number= Exon_2
FT 2576..2853
FT /*tag= f
FT /number= Intron_2
FT 2854..3107
FT /*tag= g
FT /number= Exon_3
FT 3108..4302
FT /*tag= h
FT /number= Intron_3
FT 3126..3243
FT /*tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT 3329..3396
FT /*tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT 3487..3759
FT /*tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT 3782..4070
FT /*tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT 4303..4634
FT /*tag= m
FT /number= Exon_4
FT 4635..5546
FT /*tag= n
FT intron


```

FT FT /number= Intron_4
FT FT 4688..4719
FT FT /*tag= o
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5221..5330
FT FT /*tag= p
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5547..5760
FT FT /*tag= q
FT FT /number= Exon_5
FT FT 5761..5814
FT FT /*tag= r
FT FT /number= Intron_5
FT FT 5815..5942
FT FT /*tag= s
FT FT /number= Exon_6
FT FT 5943..6297
FT FT /*tag= t
FT FT /number= Intron_6
FT FT 6062..6138
FT FT /*tag= u
FT FT /rpt_type= INVERTED
FT FT 6298..6537
FT FT /*tag= v
FT FT /number= Exon_7
FT FT 6434
FT FT /*tag= w
FT FT /note= "C>T, from allele n1129 and n1164"
FT FT 6538..7012
FT FT /*tag= x
FT FT /number= Intron_7
FT FT 6567..6625
FT FT /*tag= y
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT FT 6905..6965
FT FT /*tag= z
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 6567-6625"
FT FT 7012..7075
FT FT /*tag= aa
FT FT /number= Exon_8
FT FT
FT PN WO9325685-A.
FT PD 23-DEC-1993.
FT PF 14-JUN-1993; U05701.
FT PR 12-JUN-1992; US-897788.
FT PR 20-NOV-1992; US-979638.
FT PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT PI Horvitz HR, Shaham S, Yuan J;
FT DR WPI; 94-007542/01.
FT DR P-PSDB; R53285.
FT PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT develop agents to increase or prevent cell death in organisms
FT PS Claim 14; Fig 4; 127pp; English.
FT CC The sequences given in Q64735-45 represent mutations of the C. elegans
FT CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT CC was most abundant in embryos, but was also detected in larvae and young
FT CC adults, suggesting that ced-3 is expressed not only in cells undergoing
FT CC cell death. The four largest introns as well as sequences 5' of the
FT CC start codon contain repetitive elements, some of which have been
FT CC characterised in non-coding regions of other C. elegans genes, such
FT CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
FT CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
FT CC region that might be a transmembrane region. One region of Ced-3 is
FT CC very rich in serine. It is thought that this region is involved in
FT CC protein-protein interactions, similar to acid blobs in transcription
FT CC factors. Of the mutations which occur within the ced-3 gene, eight of
FT CC the mutations are missense mutations, two are nonsense mutations and
FT CC two are putative splicing mutations. These mutations establish the
FT CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
FT CC function is not essential for viability. The ced-3 and ced-4 gene

```

```

CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.2%; Score 992; DB 9; Length 7653;
Best Local Similarity 99.6%; Pred. No. 0.00e+00;
Matches 996; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 901 gtcgtccttggtatcctcaactgtcccggtttgttttcgggtacacacttccgtgatgc 960
  |||||
QY 901 GTCGTCTCTGGTATCCTCAACTGTCCCGGTTTGTGTTTCGGTACACTCTCCGATGC 960

Db 961 cacctgtctccgtctcaattatgttttagaaaatgtgaactgtccagatgggtgactcata 1020
  |||||
QY 961 CACCTGTCTCCGTCTCAATTATCGTTTAGAAATGTGAAGTGTCCAGATGGGTGACTCATA 1020

Db 1021 ttgctgtctacaatccactttcttctcctcgcagcttctacgagcccatataaac 1080
  |||||
QY 1021 TTGCTGCTGTCTACAACTCCACTTCTTCTTCATCGGCAGTCTTACGAGCCCATATAAAC 1080

Db 1081 ttttttccgcgaaatttgcaataaaacccggcgaacatttctccaaattgttacgcaa 1140
  |||||
QY 1081 TTTTTCCTCCGCGAAATTGCAATAAACCGGCCCAAAACTTCTCCAAATTGTACGCAA 1140

Db 1141 tatatacaatccataagaatatctctcaatgtttatgatttcttcgcagcaactttctct 1200
  |||||
QY 1141 TATATACAATCCATAAGAAATATCTTCTCAATGTTTATGATTTCTTCGCAGCACTTCTCT 1200

Db 1201 tcgtgtgtaaacatcttattttataataattttccgcataaaattccgatttttgagtatta 1260
  |||||
QY 1201 TCGTGTGCTAACATCTTATTTTATAATATTTCCCGTAAATTTCCGATTTTTCAGTATTA 1260

Db 1261 attatcgttaaaattatgataataatagcaccgcaaaactactataaaatgttaaaagctcctt 1320
  |||||
QY 1261 ATTTATCGTAAATTTATCATATAATAGCACCCGAAACTACTAAAAATGGTAAAGCTCCTTT 1320

Db 1321 taaatcggctcgacattatcgattataaggaatcacaaaattctgagaatgcgtactgcgc 1380
  |||||
QY 1321 TAAATCGGCTCGACATTATCGTATTAAAGGAATCACAAAAATCTTGAGAAATGCGTACTGCGC 1380

Db 1381 aacatatttgacggcgaataatctcgtagcgaaactacagtaattctttaaataactac 1440
  |||||
QY 1381 AACATATTGACGGCAAAATATCTCGTAGCGAAACTACAGTAATTTCTTTAAATGACTAC 1440

Db 1441 ttagcgcttggtcgatttacgggctcaatttttgaaaaataatttttttttcgaattt 1500
  |||||
QY 1441 TGTAGCGCTTGTCGATTACGGGCTCAATTTTGAAAAATAATTTTTTTTTCGAATTT 1500

Db 1501 tgataacccgtaaaatcgtcacacgcctacagtagtgcatttaaaaggattactgtagtcta 1560
  |||||
QY 1501 TGATAACCCGTAAATCGTCACACGCTACAGTAGTCAATTTAAAGGATTACTGTAGTTCTA 1560

Db 1561 gtaacgacataattttgcgcgcaaatatgactgtaataacgcattctctgaattttgtgtt 1620
  |||||
QY 1561 GGTACGAGATATTTTGC CGCCCAATATGACTGTAATACGCATTTCTCTGAATTTTGTGTT 1620

Db 1621 tcgtaataatttcacaagattttggcattccactttaaaaggcgacagattttattcca 1680
  |||||
QY 1621 TCCGTAATAATTTTCAAGATTTTGGCATTTCCACTTTAAAGGCGCACAGGATTTATTCCA 1680

Db 1681 atgggtctggcgcaagaaaaagtttgatagacttttaaaattctctctgcatatttataattc 1740
  |||||
QY 1681 ATGGGTCTCGGCACGCAAAAAGTTTGATAGACTTTTAAATTTCTCCTTCATTTTAAATTC 1740

Db 1741 aattactaaaaatttttcgtgaatttttctgttaaaatttttaaaatcagtttttctaattt 1800
  |||||
QY 1741 AATTACTAAAAATTTTCGTGAATTTTCTGTGTAAAAATTTTAAAAATCAGTTTCTAATAT 1800

Db 1801 ttccaggctgacaaacagaaacaaaaacacacacacacacacacacacacacacacacacacac 1860
  |||||
QY 1801 TTCCAGGCTGACAAACAGAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 1860

```


Db 1861 taaaaataacgatttctcattgaaattgtgttttatgtt 1900
|||||
QY 1861 TAAAAATAACGATTCTCATTTGAAATGTGTTTATGTT 1900

RESULT 10
ID Q64743 standard; DNA; 7653 BP.
AC Q64743;
DT 23-JUN-1994 (first entry)
DE ced-3 (C6485T) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /*tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /*tag= b
FT /rpt_type= INVERTED
FT /note= "inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT /*tag= c
FT /number= Exon_1
FT 2367..2429
FT /*tag= d
FT /number= Intron_1
FT 2450..2575
FT /*tag= e
FT /number= Exon_2
FT 2576..2853
FT /*tag= f
FT /number= Intron_2
FT 2854..3107
FT /*tag= g
FT /number= Exon_3
FT 3108..4302
FT /*tag= h
FT /number= Intron_3
FT 3126..3243
FT /*tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit 3329..3396
FT /*tag= j
FT /rpt_type= INVERTED
FT /note= "inverted w.r.t. repeat at 3126-3243"
FT repeat_unit 3487..3759
FT /*tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT repeat_unit 3782..4070
FT /*tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT exon 4303..4634
FT /*tag= m
FT /number= Exon_4
FT 4635..5546
FT /*tag= n
FT /number= Intron_4
FT 4688..4719
FT /*tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5221..5330
FT /*tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"

FT exon 5547..5760
FT /*tag= q
FT /number= Exon_5
FT 5761..5814
FT /*tag= r
FT /number= Intron_5
FT 5815..5942
FT /*tag= s
FT /number= Exon_6
FT 5943..6297
FT /*tag= t
FT /number= Intron_6
FT 6052..6138
FT /*tag= u
FT /rpt_type= INVERTED
FT 6298..6537
FT /*tag= v
FT /number= Exon_7
FT 6485
FT /*tag= w
FT /note= "C>T, from allele n2430"
FT 6538..7012
FT /*tag= x
FT /number= Intron_7
FT 6567..6625
FT /*tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT 6905..6965
FT /*tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT 7012..7075
FT /*tag= aa
FT /number= Exon_8
FT
FT W09325685-A.
PN 23-DEC-1993.
PD 14-JUN-1993; U05701.
PF 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.
DR P-PSDB; R53286.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127pp; English.
CC The sequences given in Q64735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations. These mutations establish the
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.2%; Score 992; DB 9; Length 7653;
Best Local Similarity 99.6%; Pred. No. 0.00e+00;
Matches 996; Conservative 0; Mismatches 4; Indels 0; Gaps 0;


```
FT FT /*tag= a
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t repeat at 1490-1614"
FT FT 1490..1614
FT FT /*tag= b
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT FT 2232..2366
FT FT /*tag= c
FT FT /number= Exon_1
FT FT 2310
FT FT /*tag= d
FT FT /note= "C>T, from allele n1040"
FT FT 2367..2429
FT FT /*tag= e
FT FT /number= Intron_1
FT FT 2430..2575
FT FT /*tag= f
FT FT /number= Exon_2
FT FT 2487
FT FT /*tag= g
FT FT /note= "G>A, from allele n718"
FT FT 2576..2853
FT FT /*tag= h
FT FT /number= Intron_2
FT FT 2854..3107
FT FT /*tag= i
FT FT /number= Exon_3
FT FT 3108..4302
FT FT /*tag= j
FT FT /number= Intron_3
FT FT 3126..3243
FT FT /*tag= k
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT FT 3329..3396
FT FT /*tag= l
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT FT 3487..3759
FT FT /*tag= m
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT FT 3782..4070
FT FT /*tag= n
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 3487-3759"
FT FT 4303..4634
FT FT /*tag= o
FT FT /number= Exon_4
FT FT 4635..5546
FT FT /*tag= p
FT FT /number= Intron_4
FT FT 4688..4719
FT FT /*tag= q
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5221..5330
FT FT /*tag= r
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5547..5760
FT FT /*tag= s
FT FT /number= Exon_5
FT FT 5757
FT FT /*tag= t
FT FT /note= "G>A, from allele n2433"
FT FT 5761..5814
FT FT /*tag= u
FT FT /number= Intron_5
FT FT 5815..5942
FT FT /*tag= v
FT FT /number= Exon_6
```

```
FT FT mutation
FT FT 5940
FT FT /*tag= w
FT FT /note= "C>T, from allele n1165"
FT FT 5943..6297
FT FT /*tag= x
FT FT /number= Intron_6
FT FT 6062..6138
FT FT /*tag= y
FT FT /rpt_type= INVERTED
FT FT 6297
FT FT /*tag= z
FT FT /note= "G>A, from allele n717"
FT FT 6298..6537
FT FT /*tag= aa
FT FT /number= Exon_7
FT FT 6322
FT FT /*tag= ab
FT FT /note= "C>T, from allele n1949"
FT FT 6372
FT FT /*tag= ac
FT FT /note= "G>A, from allele n1286"
FT FT 6434
FT FT /*tag= ad
FT FT /note= "C>T, from alleles n1129 and n1164"
FT FT 6485
FT FT /*tag= ae
FT FT /note= "C>T, from allele n2430"
FT FT 6535
FT FT /*tag= af
FT FT /note= "G>A, from allele n2426"
FT FT 6538..7012
FT FT /*tag= ag
FT FT /number= Intron_7
FT FT 6567..6625
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT FT 6905..6965
FT FT /*tag= ai
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 6567-6625"
FT FT 7013..7075
FT FT /*tag= aj
FT FT /number= Exon_8
FT FT 7020
FT FT /*tag= ak
FT FT /note= "C>T, from allele n1163"
FT FT PN
FT FT WO9325685-A.
FT FT 23-DEC-1993.
FT FT PD
FT FT 14-JUN-1993; U05701.
FT FT PF
FT FT 12-JUN-1992; US-897788.
FT FT PR
FT FT 20-NOV-1992; US-979638.
FT FT PR
FT FT (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT FT PI
FT FT Horvitz HR, Shaham S, Yuan J;
FT FT WPI; 94-007542/01.
FT FT DR
FT FT P-PSDB; R47466.
FT FT PT
FT FT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT FT develop agents to increase or prevent cell death in organisms
FT FT Claim 2; Fig 4; 127pp; English.
FT FT PS
FT FT This sequence represents the C. elegans ced-3 gene. A 2.8 kb mRNA
FT FT CC was identified as the ced-3 transcript and was most abundant in
FT FT CC embryos, but was also detected in larvae and young adults, suggesting
FT FT CC that ced-3 is expressed not only in cells undergoing cell death. The
FT FT CC four largest introns as well as sequences 5' of the start codon
FT FT CC contain repetitive elements, some of which have been characterised
FT FT CC in non-coding regions of other C. elegans genes, such as fem-1, lin-12
FT FT CC and myoD. The Ced-3 protein is 503 amino acids in length. Ced-3 is
FT FT CC highly hydrophilic with no significant hydrophobic region that might
FT FT CC be a transmembrane region. One region of Ced-3 is very rich in serine.
FT FT CC It is thought that this region is involved in protein-protein inter-
FT FT CC actions, similar to acid blobs in transcription factors. Of the
FT FT CC mutations which occur within the ced-3 gene, eight of the mutations
FT FT CC are missense mutations, two are nonsense mutations and two are putative
```



```

FT FT /*tag= p
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5547..5760
FT FT /*tag= q
FT FT /number= Exon_5
FT FT 5757
FT FT /*tag= r
FT FT /note= "G>A, from allele n2433"
FT FT 5761..5814
FT FT /*tag= s
FT FT /number= Intron_5
FT FT 5815..5942
FT FT /*tag= t
FT FT /number= Exon_6
FT FT 5943..6297
FT FT /*tag= u
FT FT /number= Intron_6
FT FT 6062..6138
FT FT /*tag= v
FT FT /rpt_type= INVERTED
FT FT 6298..6537
FT FT /*tag= w
FT FT /number= Exon_7
FT FT 6538..7012
FT FT /*tag= x
FT FT /number= Intron_7
FT FT 6567..6625
FT FT /*tag= y
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT FT 6905..6965
FT FT /*tag= z
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 6567-6625"
FT FT 7012..7075
FT FT /*tag= aa
FT FT /number= Exon_8
FT FT
FT FT WO9325685-A.
FT FT 23-DEC-1993.
FT FT 14-JUN-1993; U05701.
FT FT 12-JUN-1992; US-897788.
FT FT 20-NOV-1992; US-979638.
FT FT (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT FT Horvitz HR, Shaham S, Yuan J;
FT FT WPI; 94-007542/01.
FT FT P-PSDB; R53281.
FT FT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT FT develop agents to increase or prevent cell death in organisms
FT FT Claim 14; Fig 4; 127pp; English.
FT FT The sequences given in Q64735-45 represent mutations of the C. elegans
FT FT ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT FT was most abundant in embryos, but was also detected in larvae and young
FT FT adults, suggesting that ced-3 is expressed not only in cells undergoing
FT FT cell death. The four largest introns as well as sequences 5' of the
FT FT start codon contain repetitive elements, some of which have been
FT FT characterised in non-coding regions of other C. elegans genes, such
FT FT as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
FT FT length. Ced-3 is highly hydrophilic with no significant hydrophobic
FT FT region that might be a transmembrane region. One region of Ced-3 is
FT FT very rich in serine. It is thought that this region is involved in
FT FT protein-protein interactions, similar to acid blobs in transcription
FT FT factors. Of the mutations which occur within the ced-3 gene, eight of
FT FT the mutations are missense mutations, two are nonsense mutations and
FT FT two are putative splicing mutations. These mutations establish the
FT FT null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
FT FT function is not essential for viability. The ced-3 and ced-4 gene
FT FT products may be used to develop agents for treating conditions
FT FT characterised by cell deaths, such as myocardial infarction, stroke,
FT FT degenerative disease, traumatic brain injury, hypoxia, pathogenic
FT FT infection, aging or hair loss.
FT FT Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

```

```

Query Match 99.2%; Score 992; DB 9; Length 7653;
Best Local Similarity 99.6%; Pred. No. 0.00e+00;
Matches 996; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 901 gtcgtccttggtatcctcaactgtcccggtttgttttcggtacactcttccgtgatgc 960
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 gtcgtccttggtatcctcaactgtcccggtttgttttcggtacactcttccgtgatgc 960
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 961 cacctgtctcgtctcaattatcgcttagaataatgtgaactgtccagatgggtgactcata 1020
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 961 cacctgtctcgtctcaattatcgcttagaataatgtgaactgtccagatgggtgactcata 1020
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1021 ttgctgtctacaaatccactttttctctcctcgcgcagtcattcagcccatcataaac 1080
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1021 ttgctgtctacaaatccactttttctctcctcgcgcagtcattcagcccatcataaac 1080
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1081 ttttttttcgcgaaatttgcaataaacccggcccaaaaacttttccaaattgttacgcaa 1140
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1081 ttttttttcgcgaaatttgcaataaacccggcccaaaaacttttccaaattgttacgcaa 1140
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1141 tatatacaatccataagaatatctctcaatgtttatgattcttcgcagcactttctct 1200
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1141 tatatacaatccataagaatatctctcaatgtttatgattcttcgcagcactttctct 1200
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1201 tcgtgtcctaactcttattttataatatattccgcgttaaaattccgatttttgagtatta 1260
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1201 tcgtgtcctaactcttattttataatatattccgcgttaaaattccgatttttgagtatta 1260
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1261 attatcgtaaaattatgataaatagcaccgcaaaaactactaaaaatggtaaaagctccttt 1320
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1261 attatcgtaaaattatgataaatagcaccgcaaaaactactaaaaatggtaaaagctccttt 1320
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1321 taaatcggtcgcacattatcgattaaaggaatcacaataattctgagaatgcgtactgcgc 1380
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1321 taaatcggtcgcacattatcgattaaaggaatcacaataattctgagaatgcgtactgcgc 1380
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1381 aacataatttgacggcaaaatatctcgtagcgaaactacagtaattctttaaataactac 1440
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1381 aacataatttgacggcaaaatatctcgtagcgaaactacagtaattctttaaataactac 1440
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1441 ttagcgtctgtgtcgattacgggctcaatttttgaaaaataatttttttttcgaattt 1500
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1441 ttagcgtctgtgtcgattacgggctcaatttttgaaaaataatttttttttcgaattt 1500
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1501 tgataaacccgtataatcgtcacaacgctacagtagtgcattttaaaggattactgtactta 1560
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1501 tgataaacccgtataatcgtcacaacgctacagtagtgcattttaaaggattactgtactta 1560
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1561 gctacgacataattttgcgcgccaaatatgactgtaataacgcattctctgaaattttgtgtt 1620
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1561 gctacgacataattttgcgcgccaaatatgactgtaataacgcattctctgaaattttgtgtt 1620
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1621 tccgtaataatttcacaagatttttgccattccacttttaaaggcgacaggaattattcca 1680
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1621 tccgtaataatttcacaagatttttgccattccacttttaaaggcgacaggaattattcca 1680
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1681 atgggtctcggcgcaaaaagttttgtagactttttaaattctcctgcatattttaaattc 1740
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1681 atgggtctcggcgcaaaaagttttgtagactttttaaattctcctgcatattttaaattc 1740
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1741 aattactaaaatttcgtgaaatttttctgttaaaatttttaaaatcagtttttctaaatt 1800
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1741 aattactaaaatttcgtgaaatttttctgttaaaatttttaaaatcagtttttctaaatt 1800
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1801 ttccaggctgacaaaacagaaaacacacacacacacacacacacacacacacacacacacac 1860
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1801 ttccaggctgacaaaacagaaaacacacacacacacacacacacacacacacacacacacac 1860
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1861 taaaaataacgattttctcattgaaaattgtgttttatgtt 1900
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1861 taaaaataacgattttctcattgaaaattgtgttttatgtt 1900
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


RESULT 14
ID T38196 standard; DNA; 7653 BP.
AC T38196;
DT 17-DEC-1996 (first entry)
DE Nematode Ced-3 gene.
KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;
KW cell death; apoptosis; neural degeneration; inflammation;
KW antiinflammatory; ds.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT FT /*tag= a
FT FT /label= Repeat-1
FT FT 1490..1614
FT FT /*tag= b
FT FT /label= Repeat-2
FT FT 2167..2366
FT FT /*tag= c
FT FT /codon_start= 2232..2234
FT FT 2310
FT FT /*tag= d
FT FT /label= T(n1040)
FT FT /note= "causes L27F mutation"
FT FT 2367..2429
FT FT /*tag= e
FT FT /label= Intron-1
FT FT 2430..2575
FT FT /*tag= f
FT FT 2487
FT FT /*tag= g
FT FT /label= A(n718)
FT FT /note= "causes G65R mutation"
FT FT 2576..2853
FT FT /*tag= h
FT FT /label= Intron-2
FT FT 2854..3107
FT FT /*tag= i
FT FT 3108..4302
FT FT /*tag= j
FT FT /label= Intron-3
FT FT 3126..3243
FT FT /*tag= k
FT FT /label= Repeat-1
FT FT 3329..3396
FT FT /*tag= l
FT FT /label= Repeat-1
FT FT 3487..3759
FT FT /*tag= m
FT FT /label= Repeat-2
FT FT 3782..4070
FT FT /*tag= n
FT FT /label= Repeat-2
FT FT 4303..4633
FT FT /*tag= o
FT FT 4634..5546
FT FT /*tag= p
FT FT /label= Intron-4
FT FT 4688..4719
FT FT /*tag= q
FT FT /label= Repeat-3
FT FT 5221..5330
FT FT /*tag= r
FT FT /label= Repeat-3
FT FT 5546..5760
FT FT /*tag= s
FT FT 5757
FT FT /*tag= t
FT FT /label= A(n2433)
FT FT /note= "causes G360S mutation"
FT FT 5761..5814
FT FT /*tag= u
FT FT /label= Intron-5
FT FT 5815..5942
FT FT exon

FT FT /*tag= v
FT FT 5940
FT FT /*tag= w
FT FT /label= T(n1165)
FT FT /note= "creates premature stop codon at 403"
FT FT 5943..6297
FT FT /*tag= x
FT FT /label= Intron-6
FT FT 6062..6138
FT FT /*tag= y
FT FT /label= Repeat-4
FT FT 6298..6537
FT FT /*tag= z
FT FT 6322
FT FT /*tag= aa
FT FT /label= T(n1949)
FT FT /note= "creates premature stop codon at 412"
FT FT 6372
FT FT /*tag= ab
FT FT /label= A(n1286)
FT FT 6434
FT FT /*tag= ac
FT FT /label= T(n1129,n1164)
FT FT /note= "causes A449V mutation"
FT FT 6485
FT FT /*tag= ad
FT FT /label= T(n2430)
FT FT /note= "causes A466V mutation"
FT FT 6535
FT FT /*tag= ae
FT FT /label= A(n2426)
FT FT /note= "causes E483K mutation"
FT FT 6538..7012
FT FT /*tag= af
FT FT /label= Intron-7
FT FT 6567..6625
FT FT /*tag= ag
FT FT /label= Repeat-5
FT FT 6905..6965
FT FT /*tag= ah
FT FT /label= Repeat-5
FT FT 7013..7653
FT FT /*tag= ai
FT FT /note= "in-frame stop codon at 7073-7075"
FT FT 7020
FT FT /*tag= aj
FT FT /label= T(n1163)
FT FT /note= "causes S486F mutation"
FT FT
FT FT W09625946-A1.
FT FT 29-AUG-1996.
FT FT 23-FEB-1996; U02473.
FT FT 24-FEB-1995; US-394189.
FT FT (MASI) MASSACHUSETTS INST TECHNOLOGY.
FT FT Horvitz HR, Shaham S, Yuan J;
FT FT WPI; 96-425082/42.
FT FT P-PSDB; R98754.
FT FT Ced-3 and human interleukin 1-beta convertase genes and proteins
FT FT useful to treat inflammation and diseases characterised by cell
FT FT death
FT FT Claim 18; Fig 3; 139pp; English.
FT FT The Caenorhabditis elegans ced-3 gene (T38196) was cloned by
FT FT mapping RFLPs and chromosome walking, and genomic DNA cloned
FT FT in plasmid pJ107 was sequenced. EMS-induced alleles were also
FT FT sequenced. The gene codes for a cell death protein (R98754) that
FT FT is structurally similar to human interleukin-1 beta converting
FT FT enzyme (ICE) (R98755), suggesting that Ced-3 protein may be a
FT FT cysteine protease like ICE and that ICE may be a human equivalent
FT FT of the nematode cell death gene. The ced-3 gene can be used as a
FT FT probe or in the prodn. of Ced-3 protein and novel drugs for
FT FT enhancing or inhibiting the activity of ICE, ced-3 and related
FT FT genes for the treatment of inflammatory diseases and/or diseases
FT FT caused by cell death. Novel inhibitors of ced-3 activity include
FT FT portions of the ced-3 gene and its product.


```

SQ Sequence 7653 BP; 2429 A; 1456 C; 1270 G; 2498 T;
Query Match 14.5%; Score 145; DB 24; Length 7653;
Best Local Similarity 83.3%; Pred. No. 1.27e-62;
Matches 244; Conservative 0; Mismatches 39; Indels 10; Gaps 8;

Db 1342 tattaaggaatcacaaaattctgagaatcgctactgacgcaacatatatttgacg-gcaaaaat 1400
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1629 TATTACGGAACACAAAATTCAGAGAATGCGTATTACAGT-CATATTTGGCGCGCAAAAT 1571
Db 1401 atctcgtagcgaaactacagtaattctttaaataactgactactgtagcgcttggtcgattt 1460
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1570 ATCTCGTAGCTAGAACTACAGTAATCCTTTAAATGACTACTGTAGCG-TTGTGACGATTT 1512
Db 1461 acgggct--caatttttgaataataatttttttttgcgaattttgataacccgtaaatcgt 1518
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1511 ACGGGTTATCAAAATTCGAAAAAATAAATTTATTTCAAAAATTGAG--CCCGTAAATCGA 1454
Db 1519 cacaac-cgctacagtagtcatttaaaggattactgttagttcttagctacgagatatatttgc 1577
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1453 CACAAGCGCTACAGTAGTCATTTAAAGAAATTACTGTAGTTTCGCTACGAGATATTTTGC 1394
Db 1578 gcgcaaatatgact-gtaatacgcattctctgaattttgtgtttccgtaata 1629
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 1393 -CGTCAATATGTTGGCAGTACGCATCTCTCAGAAATTTTGTGATTCCTTAATA 1342

RESULT 15
ID Q64744 standard; DNA; 7653 BP.
AC Q64744;
DT 23-JUN-1994 (first entry)
UN ced-3 (G6536A) gene.
AW C. elegans; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /*tag= a
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat at 1490-1614"
FT /*tag= b
FT /*rpt_type= INVERTED
FT /*note= "inverted w.r.t. repeat at 1356-1472"
FT /*tag= c
FT /*number= Exon_1
FT /*tag= d
FT /*number= Intron_1
FT /*tag= e
FT /*number= Exon_2
FT /*tag= f
FT /*number= Intron_2
FT /*tag= g
FT /*number= Exon_3
FT /*tag= h
FT /*number= Intron_3
FT /*tag= i
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat at 3329-3396"
FT /*tag= j
FT /*rpt_type= INVERTED
FT /*note= "inverted w.r.t. repeat at 3126-3243"

repeat_unit 1490..1614
exon 2367..2429
intron 2367..2429
exon 2450..2575
intron 2576..2853
exon 2854..3107
intron 3108..4302
repeat_unit 3126..3243
repeat_unit 3329..3396
repeat_unit 3487..3759

```

```

FT /*tag= k
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat at 3782-4070"
FT 3782..4070
FT /*tag= l
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat 3487-3759"
FT 4303..4634
FT /*tag= m
FT /*number= Exon_4
FT 4635..5546
FT /*tag= n
FT /*number= Intron_4
FT 4688..4719
FT /*tag= o
FT /*rpt_type= INVERTED
FT /*note= "Repeat 3"
FT 5221..5330
FT /*tag= p
FT /*rpt_type= INVERTED
FT /*note= "Repeat 3"
FT 5547..5760
FT /*tag= q
FT /*number= Exon_5
FT 5761..5814
FT /*tag= r
FT /*number= Intron_5
FT 5815..5942
FT /*tag= s
FT /*number= Exon_6
FT 5943..6297
FT /*tag= t
FT /*number= Intron_6
FT 6062..6138
FT /*tag= u
FT /*rpt_type= INVERTED
FT 6298..6537
FT /*tag= v
FT /*number= Exon_7
FT 6538
FT /*tag= w
FT /*note= "G>A, from allele n2426"
FT 6538..7012
FT /*tag= x
FT /*number= Intron_7
FT 6567..6625
FT /*tag= y
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat at 6905-6965"
FT 6905..6965
FT /*tag= z
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat 6567-6625"
FT 7012..7075
FT /*tag= aa
FT /*number= Exon_8
FT 709325685-A.
FT 23-DEC-1993.
FT 14-JUN-1993; U05701.
FT 12-JUN-1992; US-897788.
FT 20-NOV-1992; US-979638.
FT (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT Horvitz HR, Shaham S, Yuan J;
FT WPI; 94-007542/01.
FT P-PSDB; R53287.
FT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT develop agents to increase or prevent cell death in organisms
FT Claim 14; Fig 4; 127pp; English.
FT The sequences given in Q64735-45 represent mutations of the C. elegans
FT ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT was most abundant in embryos, but was also detected in larvae and young
FT adults, suggesting that ced-3 is expressed not only in cells undergoing
FT cell death. The four largest introns as well as sequences 5' of the

```


start codon contain repetitive elements, some of which have been characterised in non-coding regions of other *C. elegans* genes, such as *fem-1*, *lin-12* and *myoD*. The Ced-3 protein is 503 amino acids in length. Ced-3 is highly hydrophilic with no significant hydrophobic region that might be a transmembrane region. One region of Ced-3 is very rich in serine. It is thought that this region is involved in protein-protein interactions, similar to acid blobs in transcription factors. Of the mutations which occur within the ced-3 gene, eight of the mutations are missense mutations, two are nonsense mutations and two are putative splicing mutations. These mutations establish the null phenotype of the ced-3 gene, confirming that ced-3, like ced-4, function is not essential for viability. The ced-3 and ced-4 gene products may be used to develop agents for treating conditions characterised by cell deaths, such as myocardial infarction, stroke, degenerative disease, traumatic brain injury, hypoxia, pathogenic infection, aging or hair loss.

Sequence	7653 BP;	2430 A;	1453 C;	1271 G;	2499 T;
Sequence	7653 BP;	2430 A;	1453 C;	1271 G;	2499 T;

Query Match 14.3%; Score 143; DB 9; Length 7653;
Best Local Similarity 82.9%;
Pred. NO. 1.64e-61;
Matches 243; Conservative 0; Mismatches 40; Indels 10; Gaps 8;

Db 1342 tattaaggaatcacaaaattctgagaatgcgtactgcgcaacatatattgacg-gcaaaat 1400

Cp 1629 TATTACGGAAACACAAATTTCAGAGAAATGCGTATTACAGT-CATATTGGCCGCCAAAAT 1571

Db 1401 atctcgtagcgaaaactacagtaattctttaaataatgactactgtagcgcttggtcgattt 1460

Cp 1570 ATCTCGTAGCTAGAACTACAGTAATCCTTTAAATGACTACTGTAGCG-TTGTGACGATT 1512

Db 1461 acqqqct--caatttttqaaaaataatttttttttttcqaaattttqataaacccqtaaatcqt 1518

Cp 1511 ACGGGTTATCAAAATTCGAAAAAAAAATTTTCAAAATTTGAG--CCCGTAAATCGA 1454

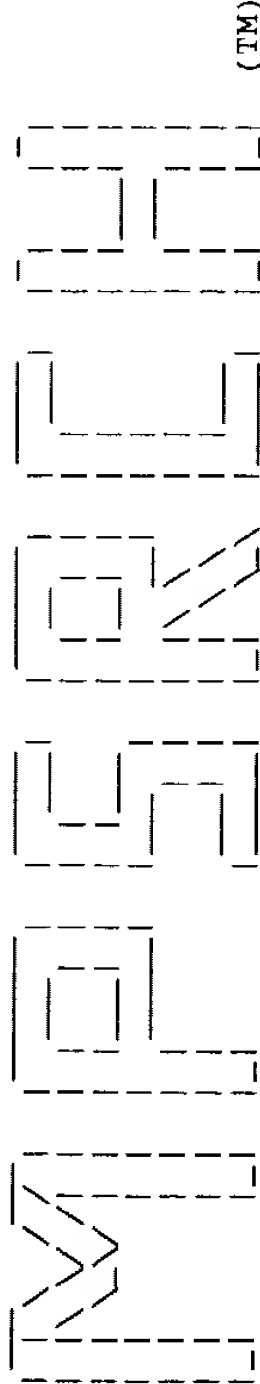
Db 1519 cacaacgctacagtagtcatttaaaggattactatagttctagctacgacatatatttcg 1577

Co 1453 CACAAGCGCTACAGTAGCTCATTTAAAGAAATTACTGTAGTTTTCCGCTACGAGATATTTTGC 1394

pb 1578 ccgccaaatattact-ctaataccatctctaaattctctccctaat 1629

CG 1393 -CGTCAAAATATGTTGGCAGTACGCATTCTCAGAAATTTTGTGATTCTTAATA 1342

Search completed: Fri Aug 6 03:10:39 1999
Job time : 620 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Fri Aug 6 02:01:37 1999; MasPar time 1697.90 Seconds
Tabular output not generated. 1380.034 Million cell updates/sec

Title: >US-08-287-669-18
Description: (901-1900) from US08287669.seq (2 of 10)
Perfect Score: 1000
N.A. Sequence: 901 GTCGTCCTTGGTATCTCAA.....TCGAAATGTGTTTATGTT 1900
Comp: CAGCAGGAACCATAGGAGTT.....ACTTTAACACAAATACAA

scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl-est58
Database: 1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
6:em_est9 7:em_gss1
genbank-est111
8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
17:gb_est18 18:gb_est19 19:gb_est20 20:gb_est21
21:gb_est22 22:gb_est23 23:gb_est24 24:gb_est25
25:gb_est26 26:gb_est27 27:gb_est28 28:gb_est29
29:gb_est30 30:gb_est31 31:gb_est32 32:gb_est33 33:gb_est34
34:gb_est35 35:gb_est36 36:gb_est37 37:gb_est38 38:gb_est39
39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 11.523; Variance 3.278; scale 3.515

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
C 1	91	9.1	359	8	T01235	WEST01956 Early embryo	5.48e-81
2	82	8.2	359	8	T01235	WEST01956 Early embryo	4.99e-69
C 3	68	6.8	300	15	C52036	C52036 Yuji Kohara unp	7.47e-51
C 4	55	5.5	360	32	D75308	CELK098H4F Yuji Kohara	1.26e-34
5	54	5.4	252	17	AA754459	97SN1787 Rice Immature	2.06e-33
C 6	54	5.4	357	15	C41501	C41501 Yuji Kohara unp	2.06e-33
7	51	5.1	337	32	D64660	CELK086DZR Yuji Kohara	8.30e-30
C 8	50	5.0	252	17	AA754459	97SN1787 Rice Immature	1.28e-28
9	50	5.0	300	15	C29898	C29898 Yuji Kohara unp	1.28e-28
10	49	4.9	357	15	C41501	C41501 Yuji Kohara unp	1.95e-27

C 11	46	4.6	247	17	AA754458	97SN1784 Rice Immature	6.24e-24
12	46	4.6	300	15	C32017	C32017 Yuji Kohara unp	6.24e-24
C 13	43	4.3	300	15	C53842	C53842 Yuji Kohara unp	1.69e-20
14	43	4.3	360	32	D75308	CELK098H4F Yuji Kohara	1.69e-20
15	42	4.2	300	15	C52036	C52036 Yuji Kohara unp	2.26e-19
16	41	4.1	247	17	AA754458	97SN1784 Rice Immature	2.96e-18
C 17	39	3.9	337	32	D64660	CELK086DZR Yuji Kohara	4.73e-16
C 18	38	3.8	300	15	C29898	C29898 Yuji Kohara unp	5.75e-15
C 19	36	3.6	300	15	C32017	C32017 Yuji Kohara unp	7.86e-13
20	35	3.5	300	15	C53842	C53842 Yuji Kohara unp	8.80e-12
21	34	3.4	300	15	C29734	C29734 Yuji Kohara unp	9.54e-11
22	34	3.4	2275	20	AF034173	AF034173 Human mRNA (T	9.54e-11
23	33	3.3	306	22	AI082072	oz52g03.x1 Soares_sene	1.00e-09
24	33	3.3	356	23	AI183776	gel7h08.x1 Soares_feta	1.00e-09
25	33	3.3	374	20	AA918212	on82g01.s1 Soares_NFL_	1.00e-09
26	33	3.3	400	8	M79744	WEST00281 Mixed stage,	1.00e-09
27	33	3.3	424	16	AA622274	no42h08.s1 NCI_CGAP_Pr	1.00e-09
28	33	3.3	525	24	AI188773	qd16f05.x1 Soares_plac	1.00e-09
C 29	32	3.2	300	15	C35709	C35709 Yuji Kohara unp	1.01e-08
30	31	3.1	215	26	AI381412	tc50f01.x1 Soares_tota	9.88e-08
31	31	3.1	225	18	AA766012	oallc02.s1 NCI_CGAP_GC	9.88e-08
32	31	3.1	244	24	AI198459	qf88a07.x1 Soares_plac	9.88e-08
33	31	3.1	315	21	AI016448	ot78f05.s1 Soares_tota	9.88e-08
34	31	3.1	326	29	AI568828	th16f10.x1 NCI_CGAP_Pr	9.88e-08
35	31	3.1	417	23	AI149055	qc81b08.x1 Soares_plac	9.88e-08
36	31	3.1	429	25	AI299057	qn35e09.x1 NCI_CGAP_Ki	9.88e-08
37	31	3.1	429	22	AI093094	qa97a10.x1 Soares_feta	9.88e-08
38	31	3.1	429	23	AI151518	qc80b12.x1 Soares_plac	9.88e-08
39	31	3.1	440	20	AA907140	qd44c05.s1 NCI_CGAP_GC	9.88e-08
40	31	3.1	441	23	AI138427	qd84d07.x1 Soares_test	9.88e-08
41	31	3.1	477	18	AA804752	ob98e04.s1 NCI_CGAP_GC	9.88e-08
42	31	3.1	494	35	AA039776	zf10c10.s1 Soares_feta	9.88e-08
43	31	3.1	505	17	AA722943	zg85a06.s1 Soares_feta	9.88e-08
C 44	31	3.1	2275	20	AF034173	AF034173 Human mRNA (T	9.88e-08
45	30	3.0	563	23	AI146823	qb92f01.x1 Soares_feta	9.26e-07

ALIGNMENTS

RESULT 1
LOCUS T01235 359 bp mRNA EST 10-NOV-1992
DEFINITION WEST01956 Early embryo, Stratagene (cat. #937007) Caenorhabditis elegans cDNA clone CEESO41, mRNA sequence.
ACCESSION T01235
NID g277716
VERSION T01235.1 GI:277716
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 359)
AUTHORS McCombie,W.R., Kelley,J.M., Aubin,L., Goscoechea,M., FitzGerald,M.G., Wu,A., Adams,M.D., Dubnick,M., Kerlavage,A.R., Venter,J.C. and Fields,C.A.
TITLE Caenorhabditis elegans cDNAs
JOURNAL Unpublished (1993)
COMMENT Other_ESTs: WEST01957
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
Seq primer: M13 Forward.
Location/Qualifiers
1. .359
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="CEESO41"
/clone_lib="Early embryo, Stratagene (cat. #937007)"
BASE COUNT 115 a 61 c 60 g 122 t 1 others


```
ORIGIN
Query Match          9.1%; Score 91; DB 8; Length 359;
Best Local Similarity 86.3%; Pred. No. 5.48e-81;
Matches 113; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

Db 1 CCCGTAATCGACACAGCCNTACAGTAGTCATTTAAAGGATTACTGTAGTTTCGCTAT 60
|||||
Cp 1465 CCCGTAATCGACACAGCGCTACAGTAGTCATTTAAAGAATTACTGTAGTTTCGCTAC 1406
|||||

Db 61 GAGATATTTGCGCGTGAATATGTTGCAATACGTAATCTCTGAATTTTGCGTTACG 120
|||||
Cp 1405 GAGATATTTGCGCGTGAATATGTTGCGCAGTACGCAATCTCAGAATTTTGTGATTCCT 1347
|||||

Db 121 TAATACTAANA 131
|||||
Cp 1346 TAATACGATAA 1336

RESULT 2
LOCUS T01235 359 bp mRNA EST 10-NOV-1992
DEFINITION wEST01956 Early embryo, Stratagene (cat. #937007) Caenorhabditis
elegans CDNA clone CERS041, mRNA sequence.
ACCESSION T01235
NID 9277716
VERSION T01235.1 GI:277716
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 359)
AUTHORS McCombie,W.R., Kelley,J.M., Aubin,L., Goscoechea,M.,
FitzGerald,M.G., Wu,A., Adams,M.D., Dubnick,M., Kerlavage,A.R.,
Venter,J.C. and Fields,C.A.
Caenorhabditis elegans cDNAs
Unpublished (1993)
Other_ESTs: wEST01957
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
Seq primer: M13 Foward.
FEATURES
source
Location/Qualifiers
1..359
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="CERS041"
/clone_lib="Early embryo, Stratagene (cat. #937007)"
BASE COUNT 115 a 61 c 60 g 122 t 1 others
ORIGIN
Query Match          8.2%; Score 82; DB 8; Length 359;
Best Local Similarity 86.4%; Pred. No. 4.99e-69;
Matches 108; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

Db 1 CCCGTAATCGACACAGCCNTACAGTAGTCATTTAAAGGATTACTGTAGTTTCGCTAT 60
|||||
QY 1507 CCCGTAATCGTCACAA-CGCTACAGTAGTCATTTAAAGGATTACTGTAGTTCTAGCTAC 1565
|||||

Db 61 GAGATATTTGCGCGTGAATATGTTGCAATACGTAATCTCTGAATTTTGCGTTACG 120
|||||
QY 1566 GAGATATTTGCGCGCCAAATATGAC-TGTAATACGCATCTCTCTGAATTTTGTGTTCCG 1624
|||||

Db 121 TAATA 125
|||||
QY 1625 TAATA 1629

RESULT 3
```

```
LOCUS C52036 300 bp mRNA EST 11-SEP-1997
DEFINITION C52036 Yuji Kohara unpublished cDNA Caenorhabditis elegans CDNA
clone yk225g1 3', mRNA sequence.
ACCESSION C52036
NID 92389793
VERSION C52036.1 GI:2389793
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 300)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
Expression map of the C.elegans genome
Unpublished (1996)
On Nov 29, 1993 this sequence version replaced gi:430548.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.j.
FEATURES
source
Location/Qualifiers
1..300
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
/db_xref="taxon:6239"
/map="21"
/clone="yk225g1"
/clone_lib="Yuji Kohara unpublished cDNA"
BASE COUNT 115 a 44 c 52 g 87 t 2 others
ORIGIN
Query Match          6.8%; Score 68; DB 15; Length 300;
Best Local Similarity 91.5%; Pred. No. 7.47e-51;
Matches 75; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 178 TTTTCAAAAATCGAGCCCGTAATTCGACACAGCTCTACAGTAAGTAATTAAGAATTAC 237
|||||
Cp 1480 TTTTCAAAAATTCGAGCCCGTAATTCGACACAGCGCTACAGTAGTCATTTAAAGAATTAC 1421
|||||

Db 238 TGTAGTTTTCGCTACGAGATAT 259
|||||
Cp 1420 TGTAGTTTTCGCTACGAGATAT 1399
|||||

RESULT 4
LOCUS D75308 360 bp mRNA EST 14-DEC-1995
DEFINITION CELK098H4F Yuji Kohara unpublished cDNA Caenorhabditis elegans CDNA
clone yk98h4 5', mRNA sequence.
ACCESSION D75308
NID g1121092
VERSION D75308.1 GI:1121092
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
Tabara,H.
Toward an expression map of the C.elegans genome
Unpublished (1994)
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
```


Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@ddb.j.nig.ac.j.
Location/Qualifiers

FEATURES

source

1. .300
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(el489)"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
/db_xref="taxon:6239"
/clone="yk228a6"
/clone_lib="Yuji Kohara unpublished cDNA"
111 a 42 c 50 g 86 t 1 others

BASE COUNT

ORIGIN

Query Match 5.0%; Score 50; DB 15; Length 300;
Best Local Similarity 86.6%; Pred. No. 1.28e-28;
Matches 71; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

Db 97 GAAATGCGTATTGTCAGCATATTGACGAGCAAAATATTTGGTCGCGAAACTACAG- 155

QY 1364 GAGATCGGTACTCGCAACATATTTGACG-GCAAAATATCTCGTAGCGAAACTACAGT 1422

Db 156 AATTCTCCAAATGACTACTGTA 177

QY 1423 AATTCTTTAAATGACTACTGTA 1444

RESULT 10

LOCUS

EST 10-SEP-1997

DEFINITION C41501 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk268f11 5', mRNA sequence.

ACCESSION

NID 92377738

VERSION C41501.1 GI:2377738

KEYWORDS EST.

SOURCE Caenorhabditis elegans.

ORGANISM

REFERENCE 1 (bases 1 to 357)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.

TITLE Expression map of the C.elegans genome

JOURNAL Unpublished (1996)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1402115.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@ddb.j.nig.ac.j.
Location/Qualifiers

FEATURES

source

1. .357
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(el489)"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
/db_xref="taxon:6239"
/clone="yk268f11"
/clone_lib="Yuji Kohara unpublished cDNA"
95 a 90 c 49 g 123 t

BASE COUNT

ORIGIN

Query Match 4.9%; Score 49; DB 15; Length 357;
Best Local Similarity 86.5%; Pred. No. 1.95e-27;
Matches 64; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Db 2 CCCGTGGATCGACATCATCGCTACAGTAGTCATTAAAGGGTTACTGTAGTTTCGCTGC 61
QY 1507 CCCGTAATCGTCA-CAACGCTACAGTAGTCATTAAAGGATTACTGTAGTTCTAGCTAC 1565

Db 62 GAGATATTGGCGC 75

QY 1566 GAGATATTGGCGC 1579

RESULT 11

LOCUS

EST 20-JAN-1998

DEFINITION 97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
cDNA clone 97SN1784, mRNA sequence.

ACCESSION

NID 92801164

VERSION AA754458.1 GI:2801164

KEYWORDS EST.

SOURCE Oryza sativa.

ORGANISM

Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.

REFERENCE

AUTHORS

1 (bases 1 to 247)
Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
Lee,M.C. and Eun,M.Y.

TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed

JOURNAL Unpublished (1998)

COMMENT On Jan 14, 1998 this sequence version replaced gi:1797455.

Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers

FEATURES

source

1. .247
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5'end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1784"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
7 a 16 c 21 g 34 t 169 others

BASE COUNT

ORIGIN

Query Match 4.6%; Score 46; DB 17; Length 247;
Best Local Similarity 12.2%; Pred. No. 6.24e-24;
Matches 27; Conservative 102; Mismatches 90; Indels 2; Gaps 2;

Db 20 MNKHTHMTBBWCCVRRVGTITNNKGKNGRTTWNDCSDNAHCRYTVBWWYARSKYGYGT 79

Cp 1137 CGTAACAATTTGGAGAAAGTTTGGCCGGTTTATTGCAAAATTCGCGGAAAAAAGTT 1078

Db 80 BYYSWNVDNTGTGVGKTTVNVSNGWNNRCSNSVYVWBTAYCDY-BHYBDRANHVDDT 138

Cp 1077 TATGATGGGCTCGTAAGACTGCCGATGAGAAAGAAAGTGGATTGTAGCAGCAATAT 1018

Db 139 RCTNDRGYCNYTASDNGTSATKRVTGYDKTSDCGGGCWRKVTYGSBYBRCGVNVMVT 198

Cp 1017 GAGTACCCCATCTGGACAGTTTCACATTTCTAAACGATAATTGAGACGGAGAC-AGGTGGC 959

Db	199	TSMWTDKSTKMBSDMSRRSRVHYGRWMBNKKRGMRSNRWTD	239
Cp	958	ATCACGGGAAGAGTGTACCGAAAAACAACCGGGACAGTTG	918

RESULT	12
LOCUS	C32017 300 bp mRNA EST 09-SEP-1997
DEFINITION	C32017 Yuji Kohara unpublished cDNA Caenorhabditis elegans CDNA clone yk314a7 3', mRNA sequence.
ACCESSION	C32017
NID	92363813
VERSION	C32017.1 GI:2363813
KEYWORDS	EST.
SOURCE	Caenorhabditis elegans.
ORGANISM	Caenorhabditis elegans
	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE	1 (bases 1 to 300)
AUTHORS	Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A. and Nishigaki,A.
TITLE	Expression map of the C.elegans genome
JOURNAL	Unpublished (1996)
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1404000

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@ddbj.nig.ac.jp.

```

FEATURES
source
1. .300
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
/db_xref="taxon:6239"
/clone="yk314a7"
/clone_lib="Yuji Kohara unpublished cDNA"
103 a 47 c 58 g 82 t 10 others
BASE COUNT
CPGCTN

```

Query Match	4.6%;	Score 46;	DB 15;	Length 300;
Best Local Similarity	81.7%;	Pred. No. 6.24e-24;		
Matches	67: Conservative	0: Mismatches	13: Indels	2: Gaps
			2:	

Db
76 GAAAATGCGTATTGTGCANCATATTTGACNAGCAAAATATTTGGTCGCCGAANACCACAN - 134
|||
1364 GAGAATGCGTACTGCCCAACATATTTGACG - GC AAAATATCTCGTAGCGAAAACTACAGT 1422

nb	135	AATCCCTAAATGACTACTGTA	156
QY	1423	AATCTTTAAATGACTACTGTA	1444

RESULT	13
LOCUS	C53842 300 bp mRNA EST 11-SEP-1997
DEFINITION	C53842 Yuji Kohata unpublished cDNA Caenorhabditis elegans cdna clone yk317f5 3', mRNA sequence.
ACCESSION	C53842
NID	92391599
VERSION	C53842.1 GI:2391599
KEYWORDS	EST.

SOURCE	Caenorhabditis elegans.
ORGANISM	Caenorhabditis elegans
	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE	1 (bases 1 to 300)
AUTHORS	Kohara, Y., Mochizashi, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano, M., Miyata, A. and Nishigaki, A.
TITLE	Expression map of the C.elegans genome
JOURNAL	Unpublished (1996)

COMMENT On Apr 14, 1993 this sequence version replaced qi:716436.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.j.

FEATURES	SOURCE
1. The first two columns are labeled "FEATURES" and "SOURCE".	
2. The third column is labeled "FEATURES".	
3. The fourth column is labeled "SOURCE".	
4. The fifth column is labeled "FEATURES".	
5. The sixth column is labeled "SOURCE".	
6. The seventh column is labeled "FEATURES".	
7. The eighth column is labeled "SOURCE".	
8. The ninth column is labeled "FEATURES".	
9. The tenth column is labeled "SOURCE".	
10. The eleventh column is labeled "FEATURES".	
11. The twelfth column is labeled "SOURCE".	
12. The thirteenth column is labeled "FEATURES".	
13. The fourteenth column is labeled "SOURCE".	
14. The fifteenth column is labeled "FEATURES".	
15. The sixteenth column is labeled "SOURCE".	
16. The seventeenth column is labeled "FEATURES".	
17. The eighteenth column is labeled "SOURCE".	
18. The nineteenth column is labeled "FEATURES".	
19. The twentieth column is labeled "SOURCE".	
20. The twenty-first column is labeled "FEATURES".	
21. The twenty-second column is labeled "SOURCE".	
22. The twenty-third column is labeled "FEATURES".	
23. The twenty-fourth column is labeled "SOURCE".	
24. The twenty-fifth column is labeled "FEATURES".	
25. The twenty-sixth column is labeled "SOURCE".	
26. The twenty-seventh column is labeled "FEATURES".	
27. The twenty-eighth column is labeled "SOURCE".	
28. The twenty-ninth column is labeled "FEATURES".	
29. The thirtieth column is labeled "SOURCE".	
30. The thirty-first column is labeled "FEATURES".	
31. The thirty-second column is labeled "SOURCE".	
32. The thirty-third column is labeled "FEATURES".	
33. The thirty-fourth column is labeled "SOURCE".	
34. The thirty-fifth column is labeled "FEATURES".	
35. The thirty-sixth column is labeled "SOURCE".	
36. The thirty-seventh column is labeled "FEATURES".	
37. The thirty-eighth column is labeled "SOURCE".	
38. The thirty-ninth column is labeled "FEATURES".	
39. The fortieth column is labeled "SOURCE".	
40. The forty-first column is labeled "FEATURES".	
41. The forty-second column is labeled "SOURCE".	
42. The forty-third column is labeled "FEATURES".	
43. The forty-fourth column is labeled "SOURCE".	
44. The forty-fifth column is labeled "FEATURES".	
45. The forty-sixth column is labeled "SOURCE".	
46. The forty-seventh column is labeled "FEATURES".	
47. The forty-eighth column is labeled "SOURCE".	
48. The forty-ninth column is labeled "FEATURES".	
49. The fiftieth column is labeled "SOURCE".	
50. The fifty-first column is labeled "FEATURES".	
51. The fifty-second column is labeled "SOURCE".	
52. The fifty-third column is labeled "FEATURES".	
53. The fifty-fourth column is labeled "SOURCE".	
54. The fifty-fifth column is labeled "FEATURES".	
55. The fifty-sixth column is labeled "SOURCE".	
56. The fifty-seventh column is labeled "FEATURES".	
57. The fifty-eighth column is labeled "SOURCE".	
58. The fifty-ninth column is labeled "FEATURES".	
59. The sixtieth column is labeled "SOURCE".	
60. The sixty-first column is labeled "FEATURES".	
61. The sixty-second column is labeled "SOURCE".	
62. The sixty-third column is labeled "FEATURES".	
63. The sixty-fourth column is labeled "SOURCE".	
64. The sixty-fifth column is labeled "FEATURES".	
65. The sixty-sixth column is labeled "SOURCE".	
66. The sixty-seventh column is labeled "FEATURES".	
67. The sixty-eighth column is labeled "SOURCE".	
68. The sixty-ninth column is labeled "FEATURES".	
69. The seventieth column is labeled "SOURCE".	
70. The seventy-first column is labeled "FEATURES".	
71. The seventy-second column is labeled "SOURCE".	
72. The seventy-third column is labeled "FEATURES".	
73. The seventy-fourth column is labeled "SOURCE".	
74. The seventy-fifth column is labeled "FEATURES".	
75. The seventy-sixth column is labeled "SOURCE".	
76. The seventy-seventh column is labeled "FEATURES".	
77. The seventy-eighth column is labeled "SOURCE".	
78. The seventy-ninth column is labeled "FEATURES".	
79. The eightieth column is labeled "SOURCE".	
80. The eighty-first column is labeled "FEATURES".	
81. The eighty-second column is labeled "SOURCE".	
82. The eighty-third column is labeled "FEATURES".	
83. The eighty-fourth column is labeled "SOURCE".	
84. The eighty-fifth column is labeled "FEATURES".	
85. The eighty-sixth column is labeled "SOURCE".	
86. The eighty-seventh column is labeled "FEATURES".	
87. The eighty-eighth column is labeled "SOURCE".	
88. The eighty-ninth column is labeled "FEATURES".	
89. The ninetieth column is labeled "SOURCE".	
90. The ninety-first column is labeled "FEATURES".	
91. The ninety-second column is labeled "SOURCE".	
92. The ninety-third column is labeled "FEATURES".	
93. The ninety-fourth column is labeled "SOURCE".	
94. The ninety-fifth column is labeled "FEATURES".	
95. The ninety-sixth column is labeled "SOURCE".	
96. The ninety-seventh column is labeled "FEATURES".	
97. The ninety-eighth column is labeled "SOURCE".	
98. The ninety-ninth column is labeled "FEATURES".	
99. The hundredth column is labeled "SOURCE".	

/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
/db_xref="taxon:6239"
/map="871E01; 5; 5q21.3"
/clone="yk317f5"
/clone_lib="Yuji Kohara unpublished cDNA"

BASE COUNT	129 a	36 c	32 g	103 t
ORIGIN				
	/clone_lib="Yuj1 Konara unpublished cdna"			

Query Match 4.3%; Score 43; DB 15; Length 300;
Best Local Similarity 82.4%; Pred. NO. 1.69e-20;
Matches 61; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

[illegible]

Db 61 ATGTTACAGTAATA 74
||| ||| ||| |||
Cp 1355 GTGATTCCCTTAATA 1342

RESULT	14
LOCUS	D75308 360 bp mRNA
DEFINITION	CELK098H4F Yuji Kohara unpublished cDNA <i>Caenorhabditis elegans</i> cdna clone yk98h4 5', mRNA sequence.
	EST 14-DEC-1995

NID	g1121092
VERSION	D75308.1
KEYWORDS	EST.
SOURCE	Caenorhabditis elegans
ORGANISM	Caenorhabditis elegans

REFERENCE
Kobayashi, T., Nishigaki, A., Motohashi, T., Sugimoto, A. and Tabara, H.
1 (bases 1 to 360)
Caenorhabditis elegans; Knaabittidae; Peloderinae; Caenorhabditis.

AUTHORS
Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and Tabara, H.

TITLE
Toward an expression map of the C.elegans genome
Unpublished (1994)

JOURNAL

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240

Email: ykohara@dbj.nig.ac.j
Insert Length: 895 Std Error: 0.00
High quality sequence stop: 363.

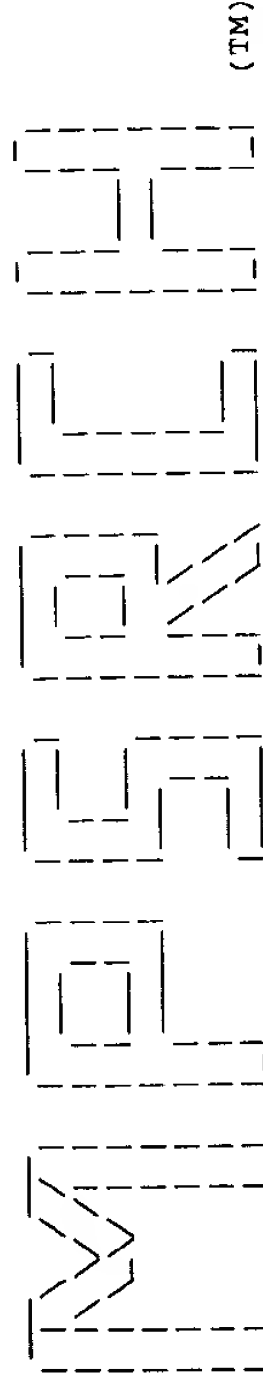
```

FEATURES
source
    "my" quarry sequence stop: 363.
    Location/Qualifiers
        1..360
            /organism="Caenorhabditis elegans"
            /strain="CB1489 him-8(e1489)"
            /note="dev_stage=varied, sex=Hermaphrodite male,
            tissue_type=whole animal"
            /db_xref="taxon:6239"
            /clone="vk98h4"

```

BASE COUNT	107 a	81 c	52 g	119 t	1 others
ORIGIN	/clone_lib="yuji kohara unpublished cdna"				

Search completed: Fri Aug 6 02:59:59 1999
Job time : 3502 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sat Aug 7 20:29:46 1999; MasPar time 1829.25 Seconds
1515.322 Million cell updates/sec
Tabular output not generated.

Title: >US-08-287-669-18
Description: (5401-6400) from US08287669.seq (7 of 10)
Perfect Score: 1000
N.A. Sequence: 5401 ATTCAAAAAAAGTCTGAAT.....TGCTCGTGGATCATGTTCA 6400
Comp: TAAGTTTTCAGCTTA.....ACGAGCACCTAGTACCAAGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb158
1:em_bal 2:em_bal 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
7:em_in 8:em_om 9:em_ov 10:em_ov 11:em_pat 12:em_ph
13:em_pl 14:em_ro 15:em_sts 16:em_vi

Database: genbank11
17:gb_bal 18:gb_bal 19:gb_htg1 20:gb_htg2 21:gb_in1
22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
28:gb_pl2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_vi

Statistics: Mean 11.316; Variance 7.880; scale 1.436

/ Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1000	100.0	7653	21	CELCED3A	Caenorhabditis elegans	0.00e+00
2	957	95.7	39908	21	CEC48D1	Caenorhabditis elegans	0.00e+00
3	955	95.5	314495	19	CEY67H2	Caenorhabditis elegans	0.00e+00
4	75	7.5	39187	21	CELC01F1	Caenorhabditis elegans	9.51e-23
5	75	7.5	45211	21	CELF35D11	Caenorhabditis elegans	9.51e-23
6	75	7.5	183422	19	AC006905	Caenorhabditis elegans	9.51e-23
7	75	7.5	331326	19	AC006802	Caenorhabditis elegans	9.51e-23
8	74	7.4	143092	21	CEY17G7B	Caenorhabditis elegans	3.27e-22
9	74	7.4	150541	21	CEY41E3	Caenorhabditis elegans	3.27e-22
10	74	7.4	207139	19	CEY17G7	Caenorhabditis elegans	3.27e-22
11	72	7.2	36554	21	CEY45F3A	Caenorhabditis elegans	3.82e-21
12	72	7.2	286715	19	CEY45F3	Caenorhabditis elegans	3.82e-21
13	69	6.9	20227	21	CEC01C7	Caenorhabditis elegans	1.49e-19

14	69	6.9	115355	19	CEY55D9	Caenorhabditis elegans	1.49e-19
15	69	6.9	224746	19	CEY56A3	Caenorhabditis elegans	1.49e-19
16	68	6.8	23961	22	CELR02C2	Caenorhabditis elegans	5.04e-19
17	68	6.8	38876	22	CELR119	Caenorhabditis elegans	5.04e-19
18	68	6.8	91286	21	CEY102A5C	Caenorhabditis elegans	5.04e-19
19	68	6.8	299202	19	AC006792	Caenorhabditis elegans	5.04e-19
20	68	6.8	299782	19	AC006844	Caenorhabditis elegans	5.04e-19
21	68	6.8	312267	19	AC006785	Caenorhabditis elegans	5.04e-19
22	67	6.7	38103	21	CELR09A1	Caenorhabditis elegans	1.69e-18
23	67	6.7	108589	19	AC006795	Caenorhabditis elegans	1.69e-18
24	67	6.7	132742	21	CEY37D8A	Caenorhabditis elegans	1.69e-18
25	67	6.7	150641	21	CEY41E3	Caenorhabditis elegans	1.69e-18
26	67	6.7	297866	19	AC006794	Caenorhabditis elegans	1.69e-18
27	66	6.6	19897	19	AC006623	Caenorhabditis elegans	5.68e-18
28	66	6.6	26195	21	CELC04G6	Caenorhabditis elegans	5.68e-18
29	66	6.6	40662	21	CEY32812B	Caenorhabditis elegans	5.68e-18
30	66	6.6	42799	22	CELF56F11	Caenorhabditis elegans	5.68e-18
31	66	6.6	125590	21	CEY49E10	Caenorhabditis elegans	5.68e-18
32	66	6.6	191857	19	CEY32B12	Caenorhabditis elegans	5.68e-18
33	66	6.6	244239	19	AC006880	Caenorhabditis elegans	5.68e-18
34	66	6.6	298406	21	CEY75B8A	Caenorhabditis elegans	5.68e-18
35	66	6.6	314495	19	CEY67H2	Caenorhabditis elegans	5.68e-18
36	66	6.6	336638	19	CEY75B8	Caenorhabditis elegans	5.68e-18
37	65	6.5	42799	22	CELF56F11	Caenorhabditis elegans	1.90e-17
38	65	6.5	48088	22	CELY44E3A	Caenorhabditis elegans	1.90e-17
39	65	6.5	244239	19	AC006880	Caenorhabditis elegans	1.90e-17
40	65	6.5	298960	20	AC006896	Caenorhabditis elegans	1.90e-17
41	65	6.5	319857	19	AC006782	Caenorhabditis elegans	1.90e-17
42	64	6.4	70594	19	AC006708	Caenorhabditis elegans	6.32e-17
43	64	6.4	110000	19	CEY111B2	Caenorhabditis elegans	6.32e-17
44	64	6.4	299670	19	AC006845	Caenorhabditis elegans	6.32e-17
45	63	6.3	224525	19	CEY70C5	Caenorhabditis elegans	2.09e-16

ALIGNMENTS

RESULT	1	CELCED3A	7653 bp	DNA	INV	23-FEB-1994
LOCUS		Caenorhabditis elegans cell death protein (ced-3)				gene, complete
DEFINITION		cds.				
ACCESSION		L29052				
NID		9456416				
VERSION		L29052.1	GI:456416			
KEYWORDS		cell death protein; interleukin-1 beta converting enzyme.				
SOURCE		Caenorhabditis elegans (strain N2)				DNA.
ORGANISM		Caenorhabditis elegans				
REFERENCE		1 (bases 1 to 7653)				
AUTHORS		Yuan, J., Shaham, S., Ledoux, S., Ellis, H.M. and Horvitz, H.				
TITLE		The C. elegans cell death gene ced-3 encodes a protein similar to				
JOURNAL		mammalian interleukin-1b-converting enzyme				
MEDLINE		Cell 75, 641-652 (1993)				
FEATURES		94061982				
source		Location/Qualifiers				
		1. .7653				
		/organism="Caenorhabditis elegans"				
		/strain="N2"				
		/db_xref="taxon:6239"				
exon		2167. .2366				
		/gene="ced-3"				
		/number=1				
gene		join(2232. .2366,2430. .2575,2854. .3107,4303. .4634,				
		5547. .5760,5815. .5942,6298. .6537,7013. .7075)				
		/gene="ced-3"				
CDS		join(2232. .2366,2430. .2575,2854. .3107,4303. .4634,				
		5547. .5760,5815. .5942,6298. .6537,7013. .7075)				
		/gene="ced-3"				
		/note="codes for a protein similar to mammalian				
		interleukin-1b-converting enzyme"				
		/codon_start=1				
		/product="cell death protein"				

/protein_id="AAA27982.1"
 /db_xref="PID:g456417"
 /db_xref="GI:456417"
 /translation="MMRQRRLLERNIMFSSHLKVDEILEVLIAKQVLSNDGDMI
 NSCGTVREKRREIVKAVQRRGDVADFADALRSTGHEGLAEVLEPLARSVDNSAVEF
 ECPWPASHRRSRALSPAGYTSPTVRHRDSVSVSSFTSYQDIYSRAFSSRSRALHS
 SDRNYSPPVNAQPSSANSSTGCSSLGYSSSRNRSFSKASGTQYIFHEEDMN
 FVDAPYISRVFDEKTMVNESSPRGMLIINNEHFEQMTNRGTKADKDLNLNLFRCM
 GYTVCKDNLTKRGMLLTIRDFAKHESHGDSAILVLHSGEENVILGVDDIPISTHEI
 YDLLNAANAPRLANKPAIVFQACGERRDNGFPVLDSDVDGPAFLFRGDNDRDGPLF
 NFLGCVRPQVQVWRKKPSQADILIRYATTAQYVSWRNSARGSWFIQAVCEVFSTHAK
 DMDVVELLTTEVNNKKVACGFQTSQGSNILKOMPENTSRLLKKFYFWPEARNsAY"
 2367. .2429

intron

BASE COUNT	2429 a	1455 c	12
ORIGIN			
	2507. .2429	/gene="ced-3"	
exon	/number=1	2430. .2575	
	/gene="ced-3"	/number=2	
intron	2576. .2853	/gene="ced-3"	
	/number=2	2854. .3107	
exon	/gene="ced-3"	/number=3	
	3108. .4302	/gene="ced-3"	
intron	/number=3	4303. .4634	
	/gene="ced-3"	/number=4	
exon	4635. .5546	/gene="ced-3"	
	/number=4	5547. .5760	
intron	/gene="ced-3"	/number=5	
	5761. .5814	/gene="ced-3"	
exon	/number=5	5815. .5942	
	/gene="ced-3"	/number=6	
intron	5943. .6297	/gene="ced-3"	
	/number=6	6298. .6537	
exon	/gene="ced-3"	/number=7	
	6538. .7012	/gene="ced-3"	
intron	/number=7	7013. .7652	
	/gene="ced-3"	/number=8	
exon	7653. .7912	/gene="ced-3"	
	/number=8	7913. .8152	
intron	/gene="ced-3"	8153. .8492	
	/number=9	8493. .8832	
exon	/gene="ced-3"	8833. .9172	
	/number=10	9173. .9512	
intron	/gene="ced-3"	9513. .9852	
	/number=11	9853. .10192	
exon	/gene="ced-3"	10193. .10532	
	/number=12	10533. .10872	
intron	/gene="ced-3"	10873. .11212	
	/number=13	11213. .11552	
exon	/gene="ced-3"	11553. .11892	
	/number=14	11893. .12232	
intron	/gene="ced-3"	12233. .12572	
	/number=15	12573. .12912	
exon	/gene="ced-3"	12913. .13252	
	/number=16	13253. .13592	
intron	/gene="ced-3"	13593. .13932	
	/number=17	13933. .14272	
exon	/gene="ced-3"	14273. .14612	
	/number=18	14613. .14952	
intron	/gene="ced-3"	14953. .15292	
	/number=19	15293. .15632	
exon	/gene="ced-3"	15633. .15972	
	/number=20	15973. .16312	
intron	/gene="ced-3"	16313. .16652	
	/number=21	16653. .16992	
exon	/gene="ced-3"	16993. .17332	
	/number=22	17333. .17672	
intron	/gene="ced-3"	17673. .18012	
	/number=23	18013. .18352	
exon	/gene="ced-3"	18353. .18692	
	/number=24	18693. .19032	
intron	/gene="ced-3"	19033. .19372	
	/number=25	19373. .19712	
exon	/gene="ced-3"	19713. .20052	
	/number=26	20053. .20392	
intron	/gene="ced-3"	20393. .20732	
	/number=27	20733. .21072	
exon	/gene="ced-3"	21073. .21412	
	/number=28	21413. .21752	
intron	/gene="ced-3"	21753. .22092	
	/number=29	22093. .22432	
exon	/gene="ced-3"	22433. .22772	
	/number=30	22773. .23112	
intron	/gene="ced-3"	23113. .23452	
	/number=31	23453. .23792	
exon	/gene="ced-3"	23793. .24132	
	/number=32	24133. .24472	
intron	/gene="ced-3"	24473. .24812	
	/number=33	24813. .25152	
exon	/gene="ced-3"	25153. .25492	
	/number=34	25493. .25832	
intron	/gene="ced-3"	25833. .26172	
	/number=35	26173. .26512	
exon	/gene="ced-3"	26513. .26852	
	/number=36	26853. .27192	
intron	/gene="ced-3"	27193. .27532	
	/number=37	27533. .27872	
exon	/gene="ced-3"	27873. .28212	
	/number=38	28213. .28552	
intron	/gene="ced-3"	28553. .28892	
	/number=39	28893. .29232	
exon	/gene="ced-3"	29233. .29572	
	/number=40	29573. .29912	
intron	/gene="ced-3"	29913	

	Query Match	100.0%;	Score 1000;	DB 21;	Length 7653;
	Best Local Similarity	100.0%;	Pred. No. 0.00e+00;		
	Matches 1000;	Conservative	0;	Mismatches 0;	Indels 0; Gaps 0;
D b	5401	ATTCAAAAAAAAAAGTCGAATTCGATTTTTTTTTTGGTTTTTTGGTCCCAAAAACCAAAA	5460		
Q y	5401	ATTCAAAAAAAAAAGTCGAATTCGATTTTTTTTTTGGTTTTTTGGTCCCAAAAACCAAAA	5460		
D b	5461	AAATCAATTTTCTGCAAAATACCAAAAGAAACCCGAAAAAATTTCCAGCCTTGTTCCT	5520		
Q y	5461	AAATCAATTTTCTGCAAAATACCAAAAGAAACCCGAAAAAATTTCCAGCCTTGTTCCT	5520		
D b	5521	AATGTAACCTGATATTTAATTTCCAGGGAATGCTCCTGACAAATTCGAGACTTTGCCAAC	5580		
Q y	5521	AATGTAACCTGATATTTAATTTCCAGGGAATGCTCCTGACAAATTCGAGACTTTGCCAAC	5580		

Db	5581	ACGAATCACAGGAGATTCTGGGATACTCGTGATTCTATCACACGGAGAGAGAATGTGA	5640
QY	5581	ACGAATCACAGGAGATTCTGGGATACTCGTGATTCTATCACACGGAGAGAGAATGTGA	5640
Db	5641	TTATTGGAGTTGATGATATACCGATTAGTACACACGAGATATATGATCTTCTCAACGCGG	5700
QY	5641	TTATTGGAGTTGATGATATACCGATTAGTACACACGAGATATATGATCTTCTCAACGCGG	5700
Db	5701	CAAATGCTCCCGCTCTGGCGAATAAGCCGAAAAATCGTTTTTGTGCAGGCTGTGCGAGGCG	5760
QY	5701	CAAATGCTCCCGCTCTGGCGAATAAGCCGAAAAATCGTTTTTGTGCAGGCTGTGCGAGGCG	5760
Db	5761	GTTCGTTTTTATTTTAAATTTAAATATAAAATTTTAAATAAAATTCATTTTCAGAACGTC	5820
QY	5761	GTTCGTTTTTATTTTAAATTTAAATATAAAATTTTAAATAAAATTCATTTTCAGAACGTC	5820
Db	5821	GTGACAATGGATTCCCAGTCTTGGATTCTGTGACGGAGTTCCTGCAATTTCTTCGTCGTG	5880
QY	5821	GTGACAATGGATTCCCAGTCTTGGATTCTGTGACGGAGTTCCTGCAATTTCTTCGTCGTG	5880
Db	5881	GATGGACAATCGAGACGGGCCATTGTTCAATTTTCTTGGATGTGTGCGGCCGCAAGTTC	5940
QY	5881	GATGGACAATCGAGACGGGCCATTGTTCAATTTTCTTGGATGTGTGCGGCCGCAAGTTC	5940
Db	5941	AGTTGCAATTTAATTTCTTGAATGAGAAATATTCCTTCAAAAAATCTAAAAATAGATTTT	6000
QY	5941	AGTTGCAATTTAATTTCTTGAATGAGAAATATTCCTTCAAAAAATCTAAAAATAGATTTT	6000
Db	6001	ATTCCAGAAAGTCCCGATCGAAAAATTCGGATATAATTACGAAATTTGTGATAAAATGAC	6060
QY	6001	ATTCCAGAAAGTCCCGATCGAAAAATTCGGATATAATTACGAAATTTGTGATAAAATGAC	6060
Db	6061	AAACCAATCAGCATCGTCGATCTCCGCCCACTTCATCGGATGGTTTGAAAGTGGCGGA	6120
QY	6061	AAACCAATCAGCATCGTCGATCTCCGCCCACTTCATCGGATGGTTTGAAAGTGGCGGA	6120
Db	6121	GTGAATTGCTGATGGTCGAGTTTTCAGTTTAGAGGGAATTTAAAAATCGCCTTTTCGA	6180
QY	6121	GTGAATTGCTGATGGTCGAGTTTTCAGTTTAGAGGGAATTTAAAAATCGCCTTTTCGA	6180
Db	6181	AAATTAATAATTGATTTTTTCAATTTTTTCGAAAAATATTCGGATTATTTTATATCTTT	6240
QY	6181	AAATTAATAATTGATTTTTTCAATTTTTTCGAAAAATATTCGGATTATTTTATATCTTT	6240
Db	6241	GGAGCGAAAGCCCCGTCCTGTAAACATTTTAAATGATAATTAATAATTTTTCAGCAA	6300
QY	6241	GGAGCGAAAGCCCCGTCCTGTAAACATTTTAAATGATAATTAATAATTTTTCAGCAA	6300
Db	6301	GTGTGGAGAAAGAACCGAGCCAAAGCTGACATTTCTGATTCGATACGCAACGACAGCTCAA	6360
QY	6301	GTGTGGAGAAAGAACCGAGCCAAAGCTGACATTTCTGATTCGATACGCAACGACAGCTCAA	6360
Db	6361	TATGTTTCGTGGAGAAACAGTCTCGTGGATCATGTTTCA	6400
QY	6361	TATGTTTCGTGGAGAAACAGTCTCGTGGATCATGTTTCA	6400

```

RESULT      2
LOCUS      CEC48D1      39908 bp      DNA      INV      23-NOV-1998
DEFINITION Caenorhabditis elegans cosmid C48D1, complete sequence.
ACCESSION  Z81049
            g1627677
VERSION    Z81049.1      GI:1627677
KEYWORDS   HTG.
SOURCE     Caenorhabditis elegans.
ORGANISM   Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE  1 (bases 1 to 39908)
AUTHORS    Burton, J.
TITLE       Direct Submission
JOURNAL     Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu

```


Cp 6160 TTCCTCTAAACTGAAACTGCGACCAATCAGCAATTCACCTCCGCCACTTTCAAACCAA 6101

Db 7263 TCCGATGAAGTGGCGGAGATCGACGATGCTGATGGTTTGTCTATTTATCACAATTC 7322

Cp 6100 TCCGATGAAGTGGCGGAGATCGACGATGCTGATGGTTTGTCTATTTATCACAATTC 6041

Db 7323 GTAATTATATCGCAATTTTTCGTCGG- ACTTCTGGAATAAAATCTATTTAGATTT 7381

Cp 6040 GTAATTATATCGCAATTTTTCGATCGGGACTTCTGGAATAAAATCTATTTAGATTT 5981

Db 7382 TTGAAGGAATATTTCTCATTTCAAGAAATTAATTCACACCTGAACTTGGCGCGCACAT 7441

Cp 5980 TTGAAGGAATATTTCTCATTTCAAGAAATTAATTCACACCTGAACTTGGCGCGCACAT 5921

Db 7442 CCAAGAAATTTGAACAATGGCCGCTCTCGATTGTCACATCCACGACGAGAAATGCAGGA 7501

Cp 5920 CCAAGAAATTTGAACAATGGCCGCTCTCGATTGTCACATCCACGACGAGAAATGCAGGA 5861

Db 7502 ACTCCGTCGACAGAAATCCAGACTGGGAATCCATTGTACAGACGTTCTGAAATGAATTT 7561

Cp 5860 ACTCCGTCGACAGAAATCCAGACTGGGAATCCATTGTACAGACGTTCTGAAATGAATTT 5801

Db 7562 ATTTAAATATTTATATTAATAATAAAACGACCGCTCGACACCTGCACA 7621

Cp 5800 ATTTAAATATTTATATTAATAATAAAACGACCGCTCGACACCTGCACA 5741

Db 7622 AAAACGATTTTCGGCTTATTCGACAGCGGGGAGCATTTGCCGCTTGAGAAGATCATAT 7681

Cp 5740 AAAACGATTTTCGGCTTATTCGACAGCGGGGAGCATTTGCCGCTTGAGAAGATCATAT 5681

Db 7682 ATCTCGTGTACTAATCGGTATATCATCACTCCAAATCAATCAATCTCTCTCCGTGT 7741

Cp 5680 ATCTCGTGTACTAATCGGTATATCATCACTCCAAATCAATCAATCTCTCTCCGTGT 5621

Db 7742 GATAGATCAGGATATCGCAGAAATCTCCGTGTGATTCGTTTGGCAAGTCTCGAAT 7801

Cp 5620 GATAGATCAGGATATCGCAGAAATCTCCGTGTGATTCGTTTGGCAAGTCTCGAAT 5561

Db 7802 GTCAGGAGCATTCCTCGGAAATTAATATCAGTTTACATTTAGGAACAAGGCTGGAAAT 7861

Cp 5560 GTCAGGAGCATTCCTCGGAAATTAATATCAGTTTACATTTAGGAACAAGGCTGGAAAT 5501

Db 7862 TTTT-GGGTTCTTTT-GGTATTTTGACGAAATATGATTTTGGTCTTTTGGACCAA 7919

Cp 5500 TTTTGGGTTCTTTTGGTATTTTGACGAAATATGATTTTGGTCTTTTGGACCAA 5441

Db 7920 AAACCAAAAAAATCGAAATTCGACTTTTGTGAAAT 7959

Cp 5440 AAACCAAAAAAATCGAAATTCGACTTTTGTGAAAT 5401

RESULT 3 CEY67H2 314495 bp DNA HTG 04-MAR-1999

LOCUS Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from

DEFINITION clone Y67H2, WORKING DRAFT SEQUENCE.

ACCESSION AL022475

NID 94469034

VERSION AL022475.3 GI:4469034

KEYWORDS HTG; HTGS_PHASE1.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans.

REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

AUTHORS Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

TITLE 1 (bases 1 to 314495)

JOURNAL McMurray, A.

Submitted (04-MAR-1999) Nematode Sequencing Project, Sanger Centre,

Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,

Washington University, St. Louis, MO 63110, USA. E-mail:

jes@sanger.ac.uk or iw@nematoe.wustl.edu

On Mar 22, 1999 this sequence version replaced gi:4468145.

IMPORTANT: This sequence is unfinished and does not necessarily

represent the correct sequence. Work on the sequence is in progress

and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments.

* NOTE: This is a 'working draft' sequence.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

FEATURES

Location/Qualifiers

1..314495

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="IV"

/clone="Y67H2"

BASE COUNT 96542 a 54551 c 52760 g 97027 t 13615 others

ORIGIN

Query Match 95.5%; Score 955; DB 19; Length 314495;

Best Local Similarity 98.5%; Pred. No. 0.00e+00;

Matches 985; Conservative 0; Mismatches 12; Indels 3; Gaps 3;

Db 236618 TGAACCATGATCCACGAGCAGCTGTTTCTCCACGAAACATATTGAGCTGTGCTGCGTATG 236677

Cp 6400 TGAACCATGATCCACGAGCAGCTGTTTCTCCACGAAACATATTGAGCTGTGCTGCGTATG 6341

Db 236678 CAATCAGAATGTCAGCTTGGCTCGGCTTCTTCTCCACACTTGTGCAAAATTTATTAA 236737

Cp 6340 GAATCAGAATGTCAGCTTGGCTCGGCTTCTTCTCCACACTTGTGCAAAATTTATTAA 6281

Db 236738 TTATCATTTAAAAATGTTTACAGACGGGCTTTTCTCCACAAAGAAATATAAAATAATCGG 236797

Cp 6280 TTATCATTTAAAAATGTTTACAGACGGGCTTTTCTCCACAAAGAAATATAAAATAATCGG 6221

Db 236798 AATATTTTCGAAAAAATTTGAAAAATCAATTTTAAATTTTCGAAAAGGCGATTTTAAA 236857

Cp 6220 AATATTTTCGAAAAAATTTGAAAAATCAATTTTAAATTTTCGAAAAGGCGATTTTAAA 6161

Db 236858 TTCCCTCTAAACTGAAACTGCGACCAATCAGCAATTCACCTCCGCCACTTTCAAACCAA 236917

Cp 6160 TTCCCTCTAAACTGAAACTGCGACCAATCAGCAATTCACCTCCGCCACTTTCAAACCAA 6101

Db 236918 TCCGATGAAGTGGCGGAGATCGACGATGCTGATGGTTTGTCTATTTATCACAATTTT 236977

Cp 6100 TCCGATGAAGTGGCGGAGATCGACGATGCTGATGGTTTGTCTATTTATCACAATTTT 6041

Db 236978 GTAATTATATCGCAATTTTTCGTCGG- ACTTCTGGAATAAAATCTATTTAGATTTT 237036

Cp 6040 GTAATTATATCGCAATTTTTCGATCGGGACTTCTGGAATAAAATCTATTTAGATTTT 5981

Db 237037 TTGAAGGAATATTTCTCATTTCAAGAAATTAATTTGCAACATGAACTTGGCGCGCACAT 237096

Cp 5980 TTGAAGGAATATTTCTCATTTCAAGAAATTAATTTGCAACATGAACTTGGCGCGCACAT 5921

Db 237097 CCAAGAAATTTGAACAATGGCCGCTCTCGATTGTCATCCACGACGAGAAATGCAGGA 237156

Cp 5920 CCAAGAAATTTGAACAATGGCCGCTCTCGATTGTCATCCACGACGAGAAATGCAGGA 5861

Db 237157 ACTCCGTCGACAGAAATCCAGACTGGGAATCCATTGTACAGACGTTCTGAAAAATGAATTT 237216

Cp 5860 ACTCCGTCGACAGAAATCCAGACTGGGAATCCATTGTACAGACGTTCTGAAAAATGAATTT 5801

Db 237217 ATTTAAATATTTATATTAATAATAAAACGACCGCTCGACAAGCCTGCACA 237276

Cp 5800 ATTTAAATATTTATATTAATAATAAAACGACCGCTCGACAAGCCTGCACA 5741

Db 237277 AAAACGATTTTCGGCTTATTCGACAGCGGGGAGCATTTGCCGCTTGAGAGATCATAT 237336

Cp 5740 AAAACGATTTTCGGCTTATTCGACAGCGGGGAGCATTTGCCGCTTGAGAGATCATAT 5681

Db 237337 ATCTCGTGTACTAATCGGTATATCATCACTCCAAATCAATCAATCTCTCTCCGTGT 237396

Cp 5680 ATCTCGTGTACTAATCGGTATATCATCACTCCAAATCAATCAATCTCTCTCCGTGT 5621

Db 237397 GATAGAAATCACGAGTATCGCAGAAATCTCCGTGTGATTCGTGTTTGGCAAGTCTCGAATT 237456
|||||
Cp 5620 GATAGAAATCACGAGTATCGCAGAAATCTCCGTGTGATTCGTGTTTGGCAAGTCTCGAATT 5561
|||||
Db 237457 GTCAGGAGCATTCCTCGGAAATTAATAATACAGTTTACATTAGGAACAAGCTGGGAAATT 237516
|||||
Cp 5560 GTCAGGAGCATTCCTCGGAAATTAATAATACAGTTTACATTAGGAACAAGCTGGGAAATT 5501
|||||
Db 237517 TTTT-CGGTTCTTTT-GGTATTTTGACGAAATTCGACTTTTGGTGGGACCAAA 237574
|||||
Cp 5500 TTTTCGGGTTTCTTTTGGTATTTTGACGAAATTCGACTTTTGGTGGGACCAAA 5441
|||||
Db 237575 AAACCAAAAAAATCGAAATTCGACTTTTGGTGGGACCAAA 237614
|||||
Cp 5440 AAACCAAAAAAATCGAAATTCGACTTTTGGTGGGACCAAA 5401
|||||

RESULT 4 CELC01F1 39187 bp DNA INV 24-MAY-1996
LOCUS
DEFINITION Caenorhabditis elegans cosmid C01F1.
ACCESSION U58761
NID g1330391
VERSION U58761.1 GI:1330391
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans strain=Bristol N2.
Caenorhabditis elegans
Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
Rhabditidae; Caenorhabditis.
REFERENCE
AUTHORS
1 (bases 1 to 39187)
Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
Saunders,D., Showkneen,R., Smaldon,N., Smith,A., Sonhammer,E.,
Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
Wilkinson-Sproat,J. and Wohldman,P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
Nature 368 (6466), 32-38 (1994)
94150718
REFERENCE
2 (bases 1 to 39187)
Johnson,D.
AUTHORS
The sequence of C. elegans cosmid C01F1
TITLE
Unpublished (1996)
JOURNAL
REFERENCE
3 (bases 1 to 39187)
Waterston,R.
AUTHORS
Direct Submission
TITLE
Submitted (19-MAY-1996) Robert Waterston
JOURNAL
COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: rwenematode.wustl.edu and jesusanger.ac.uk
NEIGHBORING COSMID INFORMATION:
The 5' cosmid is C16A11.3' end lies in a gap followed by the cosmid
B0286. Actual start of this cosmid is at base position 1 of
CELC01F1; actual end is at 39187 of CELC01F1

NOTES:
Coding sequences below are predicted from computer analysis, using
the program Genefinder(P. Green and L. Hillier, ms in preparation).
Location/Qualifiers
1. .39187
/organism="Caenorhabditis elegans"
/strain="Bristol N2"

/db_xref="taxon:6239"
/chromosome="II"
/clone="C01F1"
10034. .12946
/gene="C01F1.4"
join(10034. .10303,11261. .11463,11974. .12103,12296. .12549,
12802. .12946)
/gene="C01F1.4"
/note="similar to the family of G-protein coupled
receptors"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB00712.1"
/db_xref="PID:g1330392"
/db_xref="GI:1330392"
/translation="MDSNVKYMFEIFIPISIIILCCVAAFLNFMVVISRLYCKMRGAS
LELTYSALSDTWTSTVIGFSLFWSYKPVVLNIPHSSYCFPLTLEAFRTGGLTGLIF
HLVALAFTHYMTIKRPFDDHKKVLPRTIYIMIFFMWATPPMALMIFYFASNSGQGYQSE
KMGIKFYENFYFRALVSLIIVFLIILITTFYIKMLQKITEVRSKTASNSQTLGASAR
GRITVTVAVLIFGTFLIGWMPASILYILTAESMPLYNKHVSITISIAVLVSIMAKT
LCNPIIYATRIPEINQFVQKLLYRVLPGRNPTIRROSELEPLKTRCSQPNHVSML"
complement(13609. .15648)
/gene="C01F1.5"
complement(join(13609. .13990,14377. .14825,15553. .15648))
/gene="C01F1.5"
/note="weak similarity to C. elegans predicted protein
K09C8.3"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB00713.1"
/db_xref="PID:g1330393"
/db_xref="GI:1330393"
/translation="MRILILVSFSIVFGLATSTAVGVKKSCKRAKAFVGTKKSANGI
DAVDFEDRTLKPKQVYFVPSLTDFEKNVLRDAFAQIGRRTCVKFEQAKKPYHAD
RWDANSHVILIRKSGKFAAYDAVVEGLVDRTILYVAQSSFTSNFNSRGLVMDQLV
RFMGLQRELYRPDAVSYYQAIAGGGIPNLGTPDYNPIQLTWPFDPESITVPLWAREKFR
LTPYCPARNADIGAGQVGLLTKWDTVKLNSMYCPEKIVDADPSRGPCVVPRAKDL
SFKRLWAYKRLAKNTKKARKAVVKKHKTV"
complement(16410. .19344)
/gene="C01F1.6"
complement(join(16410. .16543,17564. .17845,18540. .18683,
18734. .18919,19014. .19344))
/gene="C01F1.6"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB00714.1"
/db_xref="PID:g1330394"
/db_xref="GI:1330394"
/translation="MSMFFQRRRESEAEIDEEVDVLMSTDIVNAHMSTKTVGFKQA
YSGWVKHAREQMGDFPVNFYSVEGLKLTTRKRREHLTSDVKKNKSLHSLSGHT
VNDDEFVVEKTONKQNRDIFSIFSEFPKSLADQSAICSAHFPSPQVWEAELNADW
LVVFFFFFLIKNPQKNEKNRDFSLSSQLNSQNTQRFAPPTFQPIRCSGRSST
LIEPPTPKIATPTGRLPTTWEYSGAAPQMGPRQIVKTNKQFKALVGMSEFP
LSVDVLVLLLEVAVPFKHLDKLRRFCSARLPPGFPVCVEIPLLATIAAKVTFQKQFT
NDIQDKMFTIPTSYREDPTRFPD"
28921. .32703
/gene="C01F1.3"
join(28921. .29160,29555. .30439,30914. .31056,31210. .31331,
32132. .32211,32278. .32703)
/gene="C01F1.3"
/note="coded for by C. elegans cDNA yk62h9.5; coded for by
C. elegans cDNA yk62h9.3; similarity to DTDG-glucose
4,6-Dehydratases"
/codon_start=1
/protein_id="AAB00715.1"
/db_xref="PID:g1330395"
/db_xref="GI:1330395"
/translation="MVTYTPKNVITGGCGFGSNFVNIHDAWPTCNFVNIDKLILNS
DTQNAESVRNSPRYKLVLTIDIKNEAAILNVEQNEIDTVIHEADCTSTRCYNETA
AVQNNVLSFIOFLETYVYKIKRFVHISTDEYVSDLSNEQGVKFSRLVPGNPY
AATKIAGEAYVYRQYQYNLPVITARMNNIYGNQWVKVVPRIEIAKVRGEYTIQ
SGKQLRSWLFVDDASAGLKAVCEKGEHELIEYNGTYEKNVADLAKTIEEVDLQGR
AHEPPKYKSIPDRPYNDRLYLISIEKAKNDLGEWPTTSFDDGMRHTVASALKEHKHV

gene

CDS

FEATURES
source


```
/db_xref="GI:868218"
/translation="MIENREFIRKSRKFAKKKKNDNEMLETIDIEDLPRPQKRFAKOF
ERVNQERVKEIFAKNYKNHISFAVLGGVIGIYWIYMYSVKQETFLLEEIDEMAATNP
KTHGHLAKK"
10364..11111
/gene="F35D11.4"
join(10364..10425,10472..10576,10634..10703,10755..11111)
/gene="F35D11.4"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA68750.1"
/db_xref="PID:g868217"
/db_xref="GI:868217"
/translation="MNSKSLFSRIIRNYSKPEPFNNMTMSRNVIEIKAVQNLDVTRRA
VEISGKOPTILKQHDIFYESPNGRLKMRSVENGVAHTELIWIYDRSDVAGPKLSNFNK
FDVPSEVLDAKLKLSQSSMGVKGKTRTLVLHGQTRIHIDRVDDLGLDFMELEVCLS
PEETPEHGEKIAHEIRELLAVPETDLLTGAYMDMLKA"
16067..18295
/gene="F35D11.3"
join(16067..16231,16323..16405,16453..16661,16780..16912,
17089..17514,17609..17715,17771..18128,18181..18295)
/gene="F35D11.3"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA68758.1"
/db_xref="PID:g868225"
/db_xref="GI:868225"
/translation="MGLSDFSFPKTEYIFHIFPTATSTPKCCSRQNAVLNLCREDFLN
EICLRHLAMGAPATRFALANLTTVLRDLDFWDDNDPNSIFFCKSTFNIVTQCLDLPEK
VIKTVKTHLEGEDLTDIAPMIMSVREDPVYKTEGSTPLASLLVAFVNOGNDSTRYR
VFLRLTLLGVVWTEFEDVEDSLASTLLEEQFVESEHSRTVREKTARNKKIKRYLMI
GAAGVGVLIGLTGGLAAPLVAASAGMLIGGAVAGLATTAGAAVLGTTMTGVAGAGF
TGYKMKKRVGATIEEFSVETLSEGVSLSCSLVSGWIESDTPDQAFVHQRHLRHTKE
QYTLRYESNYLMELGNAIEYLMSEFAVSAIQOTLLETALAGLVSAAVAPVALMSVSSV
LDNPWNVCSRAAEVGEQLAELVLSRSHGKRPIITLIGFSLGARVIFHCLLTMSKRSES
VGLIEDVILLGAPVTASPKAEWSKVCTVYSGRVINGYCEITDNLRLFLYRTMSAQFRIAG
TGPIDNRNSKKIYNYNLSHIVSY"
complement(22722..23578)
/gene="F35D11.7"
complement(join(22722..22911,22968..23004,23092..23226,
23284..23395,23450..23578))
/gene="F35D11.7"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA68753.1"
/db_xref="PID:g868220"
/db_xref="GI:868220"
/translation="MFVFLFLTAVFAKDPQCEGFSHVKRVPTARNHHTKDWCLGV
FQKNMGNDRARSFCSYVNASTIPENEHEIQVFSEIARAHGIEAPYAVDGOISPKC
NGRINKVEIILRHRFNSTEKGD SAYMVHQADNGPIFRFAADCIVLKSGIVKANGTET
RKNWGSLEFCVGNVAVQVATDDEKPYGNDVEAVLCGRHPL"
complement(24264..25135)
/gene="F35D11.8"
complement(join(24264..24438,24490..24748,24804..24915,
24968..25135))
/gene="F35D11.8"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA68754.1"
/db_xref="PID:g868221"
/db_xref="GI:868221"
/translation="MLKAFYEVQYKMLVLLFLAVFEFSKADPRCPGNYFEKRIPT
AKNNFTKGWCLGIFNETDLGNDRARSVCIIYNASLSIPENSHELEVISKTASLRNIS
VPIALDGLSPRCKFAIYKGVKGRFNTQTEEGICNKEYELFIYDDINTDTSFIRT
KLGVSAGSNVRSVQADDDIEYRSTADCTVLEKGSYHYSNGTLKDGHTNLMDCMQLAG
SKRLPEKNVAVLCGRLPF"
complement(25973..26845)
/gene="F35D11.9"
complement(join(25973..26147,26206..26461,26551..26662,
26711..26845))
/gene="F35D11.9"
/codon_start=1
/evidence=not_experimental
```

```
/protein_id="AAA68755.1"
/db_xref="PID:g868222"
/db_xref="GI:868222"
/translation="MLVLFLFLVVPVAVSAKDPKPCPNFTFVKRRTLTKNNFTKGWCL
GIFNGTDVGNDRARSVCIIYNASLSIPENSHELEISDVIRARNISLPIALDGLSP
RKALIFRGEQSQYFDQRTTEGDCNINHLFVYDDINTDTLFIRSKLGTSGATNGGY
SSQEDGVKFRFTRDCNALTDAYYFNSNGTHFYAVDGRTELLGCFGEKDKONKHEYSSV
LCGRFPL"
complement(27207..28786)
/gene="F35D11.10"
complement(join(27207..27282,27334..27414,27459..27550,
27611..27729,28399..28510,28631..28786))
/gene="F35D11.10"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA68756.1"
/db_xref="PID:g868223"
/db_xref="GI:868223"
/translation="MRSVFLILIVLGVVEASNATEKEVVCPKGFKLYERKPTORNK
FTKNWCMKVIENEDVYDRPIARSVCLDYDAVITNVENQEELEDINALIRKNNKRVAVD
GVFNPKCRGFSWVRMRKFGGLCAKEKEDDHTDPAFILTWSGEPTSSDKYQEKGDST
EIFTIAECLSLHNLQAIKNTGNLKDFLCGEEDPKQVGEVPPFVGLCGCAPV"
29633..32937
/gene="F35D11.2"
join(29633..29679,30459..30726,30870..31068,31308..31500,
31622..31937,31990..32084,32141..32194,32265..32937)
/gene="F35D11.2"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA68749.1"
/db_xref="PID:g868216"
/db_xref="GI:868216"
/translation="METADVPGDCQVRGSHAAKSTHSLNFFLPNIYLSLFVYNREM
AGQAIYKHSLDMSLQSPQSPIDPLVRDVTLPVVLKCAVIOVFSSILSQTLVSPL
AEPVHTPRMKVFGALDENQKADRAADDIKESKIPIDSVTSGVLOSTPTIVLAKS
PRRFSGFTCGLLSWFACCLVLLCLAISEVAYHRQDQAFRLKWAELRQRMGLYE
LLSQOELDRMALEKQSPVESLPLRRSDDPKPIIDVIOQDDDDNTVKVSSSEESND
VAKPLMFNQYSNEQNEKFGFLQALMDKIRKHAESMGLKQDMQVHVVEVKPLYGKNEN
NFEDETSFEQNLADGGEYAMPHQDFGGPINDVYDNNDRLTFRDHGMRPFPGPNRE
DGPPQIYDDYSEPDQQQFQWNMPPSPPPQFGFQQQPANWGPSPRSMEVQTQWNGF
POVPVGGQQFQQQQAQWYMPYNNENQAQVNONQONQDPFKQLSNVWQQNDDWMI
KNHENQMNQAMPFQPNPAASQNADEFNVPMDVISDDAVAPPSTAISADKEKAWAPPAS
SEENVISPDASLEPINDGKFLPIRSDDKPIDSADAHLFQIDDPSSFR"
34074..34780
gene
```

...
Note: remainder of annotations omitted.

Query Match 7.5%; Score 75; DB 21; Length 45211;
Best Local Similarity 77.0%; Pred. No. 9.51e-23;
Matches 114; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

```
Db 8030 ACCAATCAGCATCGTGAACCTCCGCCACTCCCTCTGATTGGTTCAACAGTGGCGGTTTC 8089
|||||
QY 6063 ACCAATCAGCATCGTGCATCTCCGCCACTTCATCGGATTGGTTTGAAGTGGCGGAGT 6122
|||||
Db 8090 CAATCGCTGATTGGTTGCTGTCTAGTTTGGAGGGAATTTCAAAATCCCGGTTTAAAA 8149
|||
QY 6123 GAATTGCTGATTGGTCGCAGTTTTCAGTTTAGAGGGAATTTAAAAATCGCCTTTTCGAAA 6182
|||
Db 8150 TTTGGATTTTTCGATTTTTCCTGTTTTCCTGTTTTCCTGTTTTCCTGTTTTCCTGTTT 8177
|||
QY 6183 ATTAATAATT-GATTTTTCATATTTTTCCTGTTTTCCTGTTTTCCTGTTTTCCTGTTT 6209
```

```
RESULT 6
LOCUS AC006905 183422 bp DNA HTG 24-FEB-1999
DEFINITION Caenorhabditis elegans clone Y74E4, WORKING DRAFT SEQUENCE, 2
unordered pieces.
ACCESSION AC006905
NID 94263443
VERSION AC006905.1 GI:4263443
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
```


Qy 6063 ACCAATCAGCATCGTCGATCTCCGCCACCTTCATCGGATTGGTTTGAAAGTGGCGGAGT 6122

Db 25417 GAATCGCTGATTGGTCGCAATCTCATTTTCGAGGGAATTGAAAATACACGCTTT 25471

Qy 6123 GAATGCTGATTGGTCGAGTTTTCAGTTTAGAGGGAATTTAAATAATCGCCTTTT 6177

RESULT 8

LOCUS CEY17G7B 143092 bp DNA INV 30-NOV-1998

DEFINITION Caenorhabditis elegans cosmid Y17G7B, complete sequence.

ACCESSION AL023828

NID 93217816

VERSION AL023828.1 GI:3217816

KEYWORDS HTG.

SOURCE

ORGANISM Caenorhabditis elegans.

Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 143092)

Smye,R.

Direct Submission

Submitted (09-JUN-1998) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rwnematode.wustl.edu

2 (bases 1 to 143092)

Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Showkeen,R., Smaldon,N., Smith,A., Sonhammer,E., Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Sproat,J. and Wohldman,P.

2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans

Nature 368 (6466), 32-38 (1994)

94150718

Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.

For a graphical representation of this sequence and its analysis see:-

http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence &object=Y17G7B

Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

IMPORTANT: This sequence is not the entire insert of clone Y17G7B. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true right end of clone W03C9 is at 100 in this sequence. The true right end of clone Y17G7 is at 143092 in this sequence. The start of this sequence (1..100) overlaps with the end of sequence 266516.

The end of this sequence overlaps with the start of sequence AL022279.

Location/Qualifiers

1..143092

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="II"

/clone="Y17G7B"

complement(3569..4874)

/gene="Y17G7B.1"

gene

CDS

complement(join(3569..3750,3804..3945,4851..4874))

/gene="Y17G7B.1"

/note="similar to Acyl CoA binding protein"

/codon_start=1

/protein_id="CAA19448.1"

/db_xref="PID:el323259"

/db_xref="PID:g3947596"

/db_xref="GI:3947596"

/translation="MPISELEKKWTGLTFEIAAEEEMRRRLKSEPTDRERLKLALYKQA LHGDIPNEDVYVPAGDEVGRKYYAAWKSQGANSEKCRADYVAIAEEMIKKYGRIV RCKWNSEVWSVDY"

11683..19658

/gene="Y17G7B.2b"

11683..19658

/gene="Y17G7B.2c"

11683..19658

/gene="Y17G7B.2a"

join(11683..11753,11836..11928,11980..12304,13242..13462,14218..14331,15865..16127,17187..17299,18228..18428,18624..18714,19438..19658)

/gene="Y17G7B.2c"

/note="Y17G7B.2a"

/codon_start=1

/protein_id="CAA19449.1"

/db_xref="PID:el323260"

/db_xref="PID:g3947597"

/db_xref="GI:3947597"

/translation="MRSKGGRRQAAPKAPTTCVYCDGKRELGSVEVVCSTCLKWF HGRCLKEFHLSNGVPEMICYFTCKQCRPTAEDWKAADLVQMCVTVLATLSAER LKADGKLSAEHVPEFTYLSLKDDEIVPYMNNWYMLTAIKQKKEWHQNLAPTLLEKN IFVQHNDLDFALAEKNSLLGPLEHVAVKLIGKRIERENRPRHIELPPIEGPKTR GASKRRHAEAPVTGKKQKLAADYSSTAAPNGVQIDIPFSKDNRYLYLTVDPNVPEDP AWNQSSAYVIPSFHYRELLNPTVNVSSNDRAFQLSINGNSITGFEYSMARASHGV SKGTWYFEVNFDDQDDSHIRIGWSQSYASLOACVGNKFSYGWSKSGTKFHEAKGK KYHFGGFKQGDVLGCLHLFVDDKLLQIPANLPSEKYLPSVSHKGFNLISFKANYFEVQ EESADIATKLVEMPGSYIEFFHNGKSCGAYENIYAGAYYPSISIFKSATATMNLGPK FRNLPRGATGIHARADEQQHEQTLSDMLYLVSKVYNLDHPPRVKREDDDDVDKDIKEI KOEI"

join(11683..11753,11836..11928,11980..12304,13242..13462,14218..14331,14622..14673,18228..18428,18624..18714,19438..19658)

/gene="Y17G7B.2c"

/note="cDNA EST EMBL:D74713 comes from this gene; cDNA EST yk470e5.3 comes from this gene; cDNA EST yk470e5.5 comes from this gene"

/codon_start=1

/protein_id="CAA19469.1"

/db_xref="PID:el356169"

/db_xref="PID:g3947617"

/db_xref="GI:3947617"

/translation="MRSKGGRRQAAPKAPTTCVYCDGKRELGSVEVVCSTCLKWF HGRCLKEFHLSNGVPEMICYFTCKQCRPTAEDWKAADLVQMCVTVLATLSAER LKADGKLSAEHVPEFTYLSLKDDEIVPYMNNWYMLTAIKQKKEWHQNLAPTLLEKN IFVQHNDLDFALAEKNSLLGPLEHVAVKLIGKRIERENRPRHIELPPIEGPKTR GASKRRHAEAPVTGKKQKLAADYSSTAAPNGVQIDIPFSKDNRYLYLTVDPNVPEDP IFPIFLAFSLFYIEFQGDVLGCLHLFVDDKLLQIPANLPSEKYLPSVSHKGFNLISFK ANYFFEVOEESADIATKLVEMPGSYIEFFHNGKSCGAYENIYAGAYYPSISIFKSAT ATMNLGPKFRNLPRGATGIHARADEQQHEQTLSDMLYLVSKVYNLDHPPRVKREDDDD VKDIKKEIKQEI"

join(11683..11753,11836..11928,11974..12304,13242..13462,14218..14331,15865..16127,17187..17299,18228..18428,18624..18714,19438..19658)

/gene="Y17G7B.2b"

/note="cDNA EST EMBL:D35727 comes from this gene"

/codon_start=1

/protein_id="CAA19468.1"

/db_xref="PID:el323261"

/db_xref="PID:g3947616"

/db_xref="GI:3947616"

/translation="MRSKGGRRQAAPKAPTTCVYCDGKRELGSVEVVCSTCLKWF HGRCLKEFHLSNGVPEMICYFTCKQCRPTAEDWKAADLVQMCVTVLATLSA ERLKADGKLSAEHVPEFTYLSLKDDEIVPYMNNWYMLTAIKQKKEWHQNLAPTLLE KNIFVQHNDLDFALAEKNSLLGPLEHVAVKLIGKRIERENRPRHIELPPIEGPK

CDS

CDS
/gene="Y41E3.1"
complement(join(3681..3975,4674..4868,4916..5496,5827..5951,6263..6549,6620..6797,6854..7184,7981..8133))
/gene="Y41E3.1"
/note="cDNA EST yk321f3.5 comes from this gene; cDNA EST EMBL:D27450 comes from this gene; cDNA EST EMBL:C08495 comes from this gene; cDNA EST yk321f3.3 comes from this gene; cDNA EST yk349c10.5 comes from this gene; cDNA EST yk467a7.5 comes from this gene"
/codon_start=1
/protein_id="CAB09003.1"
/db_xref="PID:el350440"
/db_xref="PID:g3880887"
/db_xref="GI:3880887"
/db_xref="SPTREMBL:O62436"
/translation="MATFIFRIRQLAIIGALCVSVFVLSLRLIGYDDVSRLEQSRAE RYQEIDEIPVKVRTVHSDRDRHKGEDSSSEKHAETALQSKCNIPKLEINGSEVINFFK THAPLECKKNSIEDNWVFINDEAIIRTEKRAKAKCKIQYFSRVDDNNKYEPEPVEV LDGDRMNGSDYATISCIQGFQKWKSIILVDVNDPEPEIHERARKMRKEQADPEKKPYNVY FLGFDLSQMSFRRKLPQTQVLFLEETLGSVVLNGYINVGDGTPQAFIPILTAQTETEL PLTRKRYKNANYDDVPFIWKNYSKGVVTMAEDAFSIGTFTYRLKGFQPGQPTDHY TRTLFEEVEKLNDRNCIGSIPLHRMWLQNGRQFMKKYEDVPRFLMHQSLSHDDINL VGVEDELSAHLSHMNDGLFDDSVIVMADHGRFAKLDRTHOGLQLEERMPFFSISL PKQLRNTKEKMEQNRENAEKLTSPEFIHASLMDILNLSSELGENELNRMQDASSKR SLSVFKPIPVDRTCQAGIEPHWCTCLSWKDAMDTGEDRKLTRRIANAVVREINKEIV AERDLCAPLSLAKIIDAKKYNFLKFLFKKIQKKARFSGNFMFFPKIRKKLLPKDGL LAYKNVKDAGDFVFDLSGSTKAFAHYQLKLTTPGEAIYEVTLFIDIVRDEVKLDG SSMHVNKFGDTPHCCIIDKNYLATFCVCYDRV"
complement(34904..36011)
/gene="Y41E3.2"
complement(join(34904..34955,35014..35750,35931..36011))
/note="Y41E3.2"
/note="predicted using Genefinder; similar to CUTICLE COLLAGEN 1; cDNA EST yk223f8.5 comes from this gene"
/codon_start=1
/protein_id="CAB08999.1"
/db_xref="PID:el350436"
/db_xref="PID:g3880883"
/db_xref="GI:3880883"
/db_xref="SPTREMBL:O62432"
/translation="MDIDSKIKAYRFVGYAAVCFSAAMAVLSVCITLPMVYVNVQHVRT QMHQELYACKSAKDVNSVYSIRELPLSANRTARSYNEQCAGCCLPGPQPGTGP KPGKPGKPGAPGQGTGPRPQPCPTTPPCQPCQPGPPGPPGPPGPPGPPGPPG QGKPGDAPGEPGPGKPIGPPGPGQAGAPGEPGSAKSEPAVPVPGPPGPGQAGQQG PPGPPGNGIDGAPGAPGAKGEPGTPGEPKDGEPGKPGTGGTGPGEKPGKPYCA LDGGIFFEDGTRR"
40022..45489
/gene="Y41E3.3"
join(40022..40111,40586..40699,42346..42577,43623..43993,45170..45302,45350..45489)
/gene="Y41E3.3"
/note="predicted using Genefinder"
/codon_start=1
/protein_id="CAB09000.1"
/db_xref="PID:el350437"
/db_xref="PID:g3880884"
/db_xref="GI:3880884"
/db_xref="SPTREMBL:O62433"
/translation="MDTQGTFDNNSTYQOCMTVFALSTIVSVQIYVNVVDNIQEDALQ HLSLFVEYGRAMEQPHNFGKPFQQLVFCVRDFKNQEEYEFEGNGTDFLDNVLQTNP EQPEIKQVRELLREYFEDIQYLLPHPGYKVAERQSGFRGHVKDLRPLFREELKMWVP NLLGPHILKPKIVNGKVTCTRMIOYFKEYAASFDEGTLPQPOSILNANAKLICIEAA HEAKVYNSRGMDRSTYGRMMSEKLLLEAHIKHGITALNIFDKCPRIGAKAELEKYK RLNEAKRVTCASAMLAGDSVFLGIGLSAASGAIAGAVLTLAGVWSLGIVAIPIS LTLFFIWAYVYVTKPCFEKVSNNK"
55984..56394
/gene="Y41E3.6"
55984..56394
/gene="Y41E3.6"
/note="predicted using Genefinder"
/codon_start=1
/protein_id="CAB09001.1"

gene
/db_xref="PID:el350438"
/db_xref="PID:g3880885"
/db_xref="GI:3880885"
/db_xref="SPTREMBL:O62434"
/translation="MFPCLCTLKILTFQPKFKQIHDEMKALKAAKGQDEFAKKFVTKD AIFMGPLHEPANYTEAAKFAKSEMMAAVSKGFENITIDEITPGDVVIERCTIHAKVA SGDKTGWSLVCVWVKDGGAWKIRNSCTTFKVVPSA"
57010..57193
/gene="Y41E3.5"
join(57010..57013,57063..57193)
/gene="Y41E3.5"
/note="predicted using Genefinder"
/codon_start=1
/protein_id="CAB09002.1"
/db_xref="PID:el350439"
/db_xref="PID:g3880886"
/db_xref="GI:3880886"
/db_xref="SPTREMBL:O62435"
/translation="MSKAEEQKFHDEMALKNAGKHDEFAKKFAAKVDVHRTTPRSMHR"
complement(57889..62435)
/gene="Y41E3.4"
complement(join(57889..58003,58063..58590,59633..60279,60972..61457,61780..62115,62187..62435))
/gene="Y41E3.4"
/note="predicted using Genefinder; similar to tRNA synthetases class I (E and Q); cDNA EST EMBL:D34745 comes from this gene; cDNA EST EMBL:D27420 comes from this gene; cDNA EST EMBL:D27419 comes from this gene; cDNA EST EMBL:D27418 comes from this gene; cDNA EST EMBL:C07239 comes from this gene; cDNA EST EMBL:C10280 comes from this gene; cDNA EST EMBL:C11865 comes from this gene; cDNA EST EMBL:C11446 comes from this gene; cDNA EST yk255d8.3 comes from this gene; cDNA EST yk255d8.5 comes from this gene; cDNA EST yk268e7.5 comes from this gene; cDNA EST yk293b5.3 comes from this gene; cDNA EST yk324h5.3 comes from this gene; cDNA EST yk339g3.5 comes from this gene; cDNA EST yk405a11.3 comes from this gene; cDNA EST yk436e1.3 comes from this gene; cDNA EST yk436e1.5 comes from this gene; cDNA EST yk444f8.3 comes from this gene; cDNA EST yk444f8.5 comes from this gene; cDNA EST yk450d2.3 comes from this gene; cDNA EST yk450d2.5 comes from this gene; cDNA EST EMBL:M89275 comes from this gene"
/codon_start=1
/protein_id="CAB08998.1"
/db_xref="PID:el350435"
/db_xref="PID:g3880882"
/db_xref="GI:3880882"
/db_xref="SPTREMBL:O62431"
/translation="MATKEELLSGLSDSKVAETLKNVKLTEIGSIVKLASESGEIS KQKGTLLYQLATKLPQVAAHTPLVVKYIMNDGIKTEPOLSAATIEYLLSHTVKGIOVP DFEKSCGVGVVVTIDIEAAVTKVIQHREKIVAERYSPAGKLLGELRALLPWADGA ITKKEVDLRFLELLGPKTAEDLAPKKKEKPEGPKPSKDAATAATPGTKNQKEASPE EFADGAETMDLLRTRAHFHKVGENFKQDGYVTPKTAELLKAHVAAVGGKVTRFPP EPNGVLHGHAKAININFGYAKAMGVNCLRFDDTNPEKEEEKFFSAIEDIVHWLGYD PARVTHSSDNFQQLYLMAYVKLIQKGLAFVCHOKVEEMRGFEVQLSPWRERPIEENIQL FEDMKNKGKDFEGEATLRLKLTLEEGKVDVPVARIKYVPHHRTGNQWCIYPTYDYTHCL"
...: remainder of annotations omitted.
Query Match 7.4%; Score 74; DB 21; Length 150641;
Best Local Similarity 81.9%; Pred. No. 3.27e-22;
Matches 95; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
Db 76721 AAAACGTGATTTTCAAAATCCCAACCAACAGACAGACCTGCGACCAATCAGCAATCGCTCC 76780
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 6177 AAAGCGGATTTTAAATCCCTCTAAACTGAAAACTGCGACCAATCAGCAATTCACITCC 6118
Db 76781 GCCCACTTTCAGACCAATCAGCGCGCTGGCGGAGTTTGTACGACGCTGATTGGTT 76836
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||


```
Cp 6117 GCCCACTTCAAACCAATCCGATGAAGTGGCGGAGATCGACGATGCTGATTGGTT 6062

RESULT 10
LOCUS CEY17G7 207139 bp DNA HTG 03-DEC-1998
DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
clone Y17G7, WORKING DRAFT SEQUENCE.
ACCESSION 295049
NID 93378012
VERSION 296049.1 GI:3378012
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 207139)
Smyle, R.
Direct Submission
Submitted (03-DEC-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwenematode.wustl.edu
On Aug 4, 1998 this sequence version replaced gi:2546913.
Order of segments is not known; 800 n's separate segments.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
    source
        Location/Qualifiers
            1..207139
                /organism="Caenorhabditis elegans"
                /db_xref="taxon:6239"
                /chromosome="II"
                /clone="Y17G7"
BASE COUNT 68582 a 34583 c 35567 g 67606 t 801 others
ORIGIN

Query Match 7.4%; Score 74; DB 19; Length 207139;
Best Local Similarity 83.6%; Pred. No. 3.27e-22;
Matches 92; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Db 60904 AAATTCCTTCAAATGAGAACTGCGACCAATCAGCGATTGCGACCCCACTTTTCAAC 60963
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 6163 AAATTCCTTCTAAACTGAAACTGCGACCAATCAGCAATTCATCGCCCACTTTCAAC 6104
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60964 CAATCAGATGTAGTGGCGGTGCTCGAAGACGCTGATTGTTTGCAGTT 61013
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 6103 CAATCCGATGAAGTGGCGGAGATCGACGATGCTGATTGTTTGTCAATT 6054
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
LOCUS CEY45F3A 36654 bp DNA INV 23-NOV-1998
DEFINITION Caenorhabditis elegans cosmid Y45F3A, complete sequence.
ACCESSION AL032621
NID 93810700
VERSION AL032621.1 GI:3810700
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 36654)
Steward, C.
Direct Submission
Submitted (29-OCT-1998) Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwenematode.wustl.edu
2 (bases 1 to 36654)
```

AUTHORS

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Showkeen, R., Smaldon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinon-Sproat, J. and Wohldman, P.

TITLE

2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans

JOURNAL

Nature 368 (6466), 32-38 (1994)

MEDLINE

94150718

COMMENT

Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.

For a graphical representation of this sequence and its analysis see:-

<http://webc.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence&object=Y45F3A>

Current sequence finishing criteria for the C. elegans genome

sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

IMPORTANT: This sequence is not the entire insert of clone Y45F3A. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone Y39A1 is at 36549 in this sequence. The true right end of clone T21C12 is at 106 in this sequence. The start of this sequence (1..133) overlaps with the end of sequence Z50016.

The end of this sequence (36549..36654) overlaps with the start of sequence AL031633.

FEATURES

source

Location/Qualifiers

1..36654
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"

/chromosome="III"

/clone="Y45F3A"

3710..5950

/gene="Y45F3A.2"

join(3710..3796,3947..4030,4075..4136,4720..4841,

4896..4960,5183..5298,5836..5950)

/gene="Y45F3A.2"

/note="similar to Ras family"

/codon_start=1

/protein_id="CAA21489.1"

/db_xref="PID:el350542"

/db_xref="PID:g3881017"

/db_xref="GI:3881017"

/translation="MEDYKLVFKVVLVGNAGVKTCLVRKFTQGIFFPGQSATIGVDF

MIKTVKGVNDKIKIQIWDTAQGERFSITQSYRSHAIIVLYDVSCQPSDFDCLPEWL

GEIESYANRRVLKILVGNKVDGEREVPERIGRDFSDVNQFDYFLETSALDATNVDO

LFEQVATRLTNDMKLTDERVHQFRADATNSSSTGGPIKLIDRAQTQLNSCCTRQS"

6955..12218

/gene="Y45F3A.3"

join(6955..6980,7260..7373,8074..8458,9444..9614,

9666..10009,10263..10320,10868..11273,11869..12218)

/gene="Y45F3A.3"

/note="similar to Acyl-CoA dehydrogenase; cDNA EST

EMBL:D75425 comes from this gene; cDNA EST EMBL:M89392

comes from this gene; cDNA EST EMBL:M89303 comes from this

gene; cDNA EST EMBL:D27766 comes from this gene; cDNA EST

EMBL:D27765 comes from this gene; cDNA EST EMBL:D69766

comes from this gene; cDNA EST EMBL:D72529 comes from this

100

1111

14-00000

JOURNAL Submitted (18-JUN-1998) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu

COMMENT On Jun 4, 1998 this sequence version replaced gi:2464985. Order of segments is not known; 800 n's separate segments. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.

* NOTE: This is a 'working draft' sequence.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

FEATURES Location/Qualifiers

source 1..286715

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="III"

/clone="Y45F3"

BASE COUNT 92604 a 50390 c 51409 g 91504 t 808 others

ORIGIN

Query Match 7.2%; Score 72; DB 19; Length 286715;

Best Local Similarity 79.0%; Pred. No. 3.82e-21;

Matches 98; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Db 50770 AAAACGTGATTTTGAACCCCTCCGACATGAGAACTCGACCAATCAGCGATTCTGCTCC 50829

||||| ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||

Cp 6177 AAAAGCGGATTTTAAATTCCTCTAAACTGAAACTCGGACCAATCAGCAATTCACCTCC 6118

Db 50830 GTCCACTTTCGAACCAATCAGACGGGTGGCGGAGTTAGACACGCTGATTGGTCTGTC 50889

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Cp 6117 GCCCACTTCAACCAATCCGATGAACTGGCGGAGATCGACGATGCTGATTGGTTGTC 6058

Db 50890 AGTT 50893

|||||

Cp 6057 ATTT 6054

RESULT 13

LOCUS CEC01C7 20227 bp DNA INV 23-NOV-1998

DEFINITION Caenorhabditis elegans cosmid C01C7, complete sequence.

ACCESSION Z82257

NID g1913961

VERSION Z82257.1 GI:1913961

KEYWORDS HTG.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 20227)

Smye,R.

Direct Submission

Submitted (11-NOV-1996) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu

2 (bases 1 to 20227)

Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Showkeen,R., Smaldon,N., Smith,A., Sonnhammer,E., Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Sproat,J. and Wohldman,P.

2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans

Nature 368 (6466), 32-38 (1994)

JOURNAL

MEDLINE
COMMENT

94150718

On Mar 30, 1997 this sequence version replaced gi:1694983. Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.

For a graphical representation of this sequence and its analysis see:-

http://webc.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence &object=C01C7

Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

IMPORTANT: This sequence is not the entire insert of clone C01C7. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone C01C7 is at 16867 in sequence AL032649.

The start of this sequence (1..100) overlaps with the end of sequence 299170.

Location/Qualifiers

1..20227

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="IV"

/clone="C01C7"

5500..13983

/gene="C01C7.1"

Join(5500..5573,5686..5771,6198..6372,6992..7226,7787..8107,8614..8957,9745..10114,11018..11598,12339..12810,13160..13381,13525..13575,13642..13886,13941..13983)

/gene="C01C7.1"

/note="predicted using Genefinder; similar to Src homology domain 3, tyrosine-protein kinase (ACK subfamily); cDNA EST EMBL:D68483 comes from this gene; cDNA EST yk244h4.5 comes from this gene; cDNA EST yk446g4.5 comes from this gene"

/codon_start=1

/protein_id="CAB05120.1"

/db_xref="PID:e1343671"

/db_xref="PID:g3873851"

/db_xref="GI:3873851"

/db_xref="SPTREMBL:O45232"

/translation="MREPTAGTADATLNKLLQAADLSGYESDLRRKLLKLRNADLQYVEEDLLSVGMSRPEQKRLRKYTKMFPFSGIFGKVKAFKRAFSIDRKTNSVANQDDNDHVIPIEKITLCKELGQGFSGVQAGWKNAGSDVIQAVKCVGSDKLLATSSFLQEAAMTRMRHEHVRLYGVVLDTKIMLVSELATCGSLLECLHKPALRDSFPVHVLCDYAEQIAMGNSYLELQRLIHRDLAARNLVFSPKLVISDFGLSRSLGIDGYRSEFTPNLKLPIAWCAPECINFLKFTSKSDVWAYGVTIWEMFSYGEWPKGRSGAQILELVDRKELLTRKACPEDIYMLKETHTQVQDRPTFSDIVAKFPERRAQSVRAVVDCKD SAADHLHFKKDDLIIVISRPAQYDGYWFGSLRNGKLGLEFRPTDTVAHLGSEPPCS NGTIENGFEKEKGGKKNKAKESEERERKLLISEPVGDVHTCHVGIDGTAFGLLO LDKKAMCPTSSSPSTSRGSAQAPASHTSSSTSSVHLRETVARNGVPIKETMSLRDV GPLSRDALNRDVTSPVPAPSPQPSYQPRPPRSVSSVSGNSVQVHDOFSSL DRSRGSITPTAPPLTASAAANSLKDLTGISLISPSNNLISYMDQDDHRTSRPGAI SQSTLTALSSSRKDPPIAPRGPVAAVYARGYGLNKADTKNYDLKIDKIDTPASKSD IALCEKIEDLNRLDNTNYSIGTICDYSEDRLDLSMNRITSSSTTHQPPQSSSEARIRF MTEQEVKINEKSAREHRTEDLREERKEQKEQPEQPAESLSTRTPTQEQWSSAAQAEYKLLVECGTNLQASVSPPMSPSSRLSLDRSSISPAPPRVTPPLSVR NETISMRSQEEHMEHTVEENSPKRVHIIETKLIDGPARGMSPIQKILLVAMYSLN FPNFLDRHIPAFTTPMNSGSRAPAPTPVSPAPAGSTQKPPPCRPKTRQFPFLVID ERNLAYDNLNGFGAGARVAPPVPPKPKVSFADDDKQKEVAN"

BASE COUNT 6820 a 3457 c 3154 g 6796 t

ORIGIN

Query Match

6.9%; Score 69; DB 21; Length 20227;


```
Best Local Similarity 82.2%; Pred. No. 1.49e-19;
Matches 88; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
```

Db 19269 TTTTTCAGTTCTCCTCAAAAATGAGAACCGACCACCATCAGCGATTAGCTCCGCCAATTTT 19328
||||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Cp 6168 TTTTTAAATTCCCTCTAAACTGAAAACTGCACCAATCAGCAATTCACCTCCGCCCACTTT 6109

Db 19329 TCAACCAATCAGACGGAGTGCGCGGAGTTCGAAGACGCTGATTGGTT 19375
|||||| | | | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Cp 6108 CAAACCAATCCGATGAAGTGGCGGAGATCGACGATGCTGATTGGTT 6062

RESULT 14
LOCUS CEY55D9 115355 bp DNA HTG 03-DEC-1998
DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from clone Y55D9, WORKING DRAFT SEQUENCE.
ACCESSION AL008876
NID 93378114
VERSION AL008876.1 GI:3378114
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 115355)
AUTHORS Wallis,J.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1998) Nematode Sequencing Project, Sanger Centre,
Hinxtton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT On Aug 4, 1998 this sequence version replaced gi:2612778.
IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES Location/Qualifiers
source 1..115355
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="Y55D9"

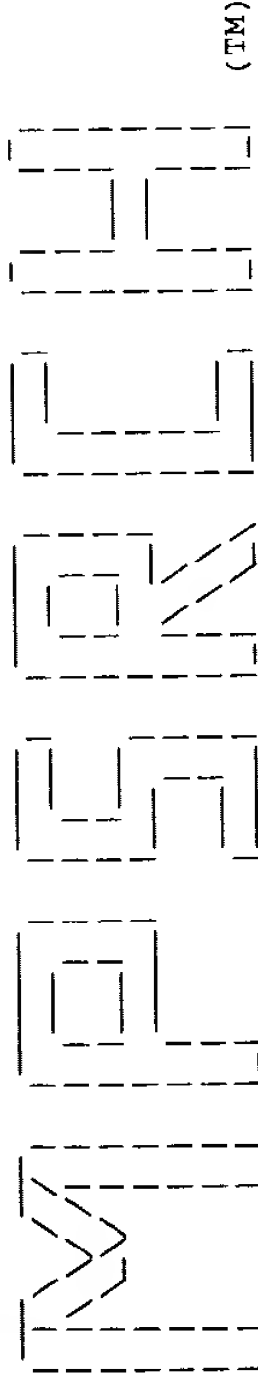
BASE COUNT 37765 a 19764 c 20592 g 37232 t 2 others
ORIGIN

Query Match 6.9%; Score 69; DB 19; Length 115355;
Best Local Similarity 82.2%; Pred. No. 1.49e-19;
Matches 88; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Db 38035 AACCAATCAGCGTCTTCGAACCTCGGCCCACTCCGTCTGATTGGTTGAAAAAATGGCGGAG 38094
|||||| | | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Qy 6062 AACCAATCAGCATCGTCGATCTCGGCCCACTTCATCGGATTGGTTTGAAAGTGGCGGAG 6121

Db 38095 CTAAATCGTGATTGGTCGCAGTCTCATTTTGGAGGAAACTGAAAAA 38141
||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Qy 6122 TGAATTGCTGATTGGTCGCAGTTTTTCAGTTTAGAGGGAATTTAAAAA 6168

RESULT 15
LOCUS CEY56A3 224746 bp DNA HTG 30-DEC-1998
DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from clone Y56A3, WORKING DRAFT SEQUENCE.
ACCESSION AL022280
NID 94090202
VERSION AL022280.1 GI:4090202
KEYWORDS HTG; HTGS_PHASE1.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sat Aug 7 21:30:12 1999; MasPar time 235.72 Seconds
909.025 Million cell updates/sec
Tabular output not generated.

Title: >US-08-287-669-18
Description: (5401-6400) from US08287669.seq (7 of 10)
Perfect Score: 1000
N.A. Sequence: 5401 ATTCAAAATAAAGTCGAAT.....TGCTCGTGATCATGTTCA 6400
Comp: TAAGTTTTTTTTCAGCTTA.....ACGAGCACCTAGTACCAAGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 9.059; Variance 6.976; scale 1.299

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	1000	100.0	7653	24	T38196	Nematode Ced-3 gene.
2	1000	100.0	7653	9	Q64742	ced-3 (C6434T) gene.
3	1000	100.0	7653	9	Q64744	ced-3 (G6536A) gene.
4	1000	100.0	7653	9	Q64743	ced-3 (C6485T) gene.
5	1000	100.0	7653	9	Q64736	ced-3 (G2487A) gene.
6	1000	100.0	7653	9	Q54666	ced-3 gene.
7	1000	100.0	7653	9	Q64745	ced-3 (C7020T) gene.
8	998	99.8	7653	9	Q54401	ced-3 gene.
9	998	99.8	7653	9	Q64737	ced-3 (G5757A) gene.

10	998	99.8	7653	9	Q64740	ced-3 (C6322T) gene.
11	998	99.8	7653	9	Q64741	ced-3 (G6372A) gene.
12	998	99.8	7653	9	Q64739	ced-3 (G6297A) gene.
13	998	99.8	7653	9	Q64738	ced-3 (C5940T) gene.
14	215	21.5	2485	14	Q79970	Interleukin-1 beta co
15	43	4.3	91	46	V44650	Mammalian DNA replica
16	43	4.3	91	9	N81164	Oligonucleotide probe
17	43	4.3	204	1	N81164	Base substituted E.co
18	43	4.3	204	1	N81164	Base substituted E.co
19	37	3.7	91	9	Q51746	Oligonucleotide probe
20	34	3.4	114	12	Q70468	Generic DNA sequence
21	34	3.4	114	12	Q70466	Generic DNA sequence
22	33	3.3	114	12	Q70465	Generic DNA sequence
23	33	3.3	114	12	Q70469	Generic DNA sequence
24	32	3.2	114	12	Q70468	Generic DNA sequence
25	32	3.2	114	12	Q70467	Generic DNA sequence
26	32	3.2	114	12	Q70471	Generic DNA sequence
27	32	3.2	114	12	Q70465	Generic DNA sequence
28	32	3.2	501	3	N50026	Sequence encoding new
29	31	3.1	114	12	Q70467	Generic DNA sequence
30	30	3.0	114	12	Q70466	Generic DNA sequence
31	30	3.0	114	12	Q70469	Generic DNA sequence
32	30	3.0	114	12	Q70472	Generic DNA sequence
33	30	3.0	7653	9	Q64744	ced-3 (G6536A) gene.
34	30	3.0	7653	9	Q64745	ced-3 (C7020T) gene.
35	30	3.0	7653	9	Q54401	ced-3 gene.
36	30	3.0	7653	9	Q64738	ced-3 (C5940T) gene.
37	30	3.0	7653	9	Q64742	ced-3 (C6434T) gene.
38	30	3.0	7653	9	Q64740	ced-3 (C6322T) gene.
39	30	3.0	7653	9	Q64737	ced-3 (G5757A) gene.
40	30	3.0	7653	9	Q54666	ced-3 gene.
41	30	3.0	7653	9	Q64736	ced-3 (G2487A) gene.
42	30	3.0	7653	9	Q64743	ced-3 (C6485T) gene.
43	30	3.0	7653	9	Q64741	ced-3 (G6372A) gene.
44	30	3.0	7653	9	Q64739	ced-3 (G6297A) gene.
45	30	3.0	7653	24	T38196	Nematode Ced-3 gene.

ALIGNMENTS

RESULT 1
ID T38196 standard; DNA; 7653 BP.
AC T38196;
DT 17-DEC-1996 (first entry)
DE Nematode Ced-3 gene.
KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;
cell death; apoptosis; neural degeneration; inflammation;
antiinflammatory; ds.
KW Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /*tag= a
FT /*label= Repeat-1
FT repeat_unit 1490..1614
FT /*tag= b
FT /*label= Repeat-2
FT exon 2167..2366
FT /*tag= c
FT /*codon_start= 2232..2234
FT 2310
FT /*tag= d
FT /*label= T(n1040)
FT /*note= "causes L27F mutation"
FT intron 2367..2429
FT /*tag= e
FT /*label= Intron-1
FT exon 2430..2575
FT /*tag= f
FT 2487
FT /*tag= g
FT /*label= A(n718)
FT /*note= "causes G55R mutation"
FT intron 2576..2853

Db	5821	gtgacaatggattccagtccttgattctgtcgacggagtgctcgtcattcttcgtcgtg	5880	FT	exon	2854..3107
QY	5821	gtgacaatggattccagtccttgattctgtcgacggagtgctcgtcattcttcgtcgtg	5880	FT	intron	3108..4302
Db	5881	gatggacaatcgagacgggccattgttcaatttcttgatgtgtgcgcgcgaagtcc	5940	FT	repeat_unit	3126..3243
QY	5881	gatggacaatcgagacgggccattgttcaatttcttgatgtgtgcgcgcgaagtcc	5940	FT	repeat_unit	3329..3396
Db	5941	aggttgcaatttaatttcttgatgagaaatctccttcacaaaatctaaaatagatttt	6000	FT	repeat_unit	3487..3759
QY	5941	aggttgcaatttaatttcttgatgagaaatctccttcacaaaatctaaaatagatttt	6000	FT	repeat_unit	3782..4070
Db	6001	attccagaaatcccgatcgacgaaatgcatataattacgaaatttgatataaaatgac	6060	FT	repeat_unit	4303..4634
QY	6001	attccagaaatcccgatcgacgaaatgcatataattacgaaatttgatataaaatgac	6060	FT	repeat_unit	4635..5546
Db	6061	aaaccaatcagcatcgatcctccgcccacttcacgattggttgaaagtggcgga	6120	FT	repeat_unit	4688..4719
QY	6061	aaaccaatcagcatcgatcctccgcccacttcacgattggttgaaagtggcgga	6120	FT	repeat_unit	5221..5330
Db	6121	gtgaattgctgattggtcgcagttttcagtttagaggggaatttaaaatcgcccttttcga	6180	FT	repeat_unit	5547..5760
QY	6121	gtgaattgctgattggtcgcagttttcagtttagaggggaatttaaaatcgcccttttcga	6180	FT	repeat_unit	5761..5814
Db	6181	aaattaaaaattgattttttcaatttttttcgaaaaaatattccgattattttatcttt	6240	FT	repeat_unit	5815..5942
QY	6181	aaattaaaaattgattttttcaatttttttcgaaaaaatattccgattattttatcttt	6240	FT	repeat_unit	5943..6297
Db	6241	ggagcgaagcccgctcgtgtaaacatttttaaatgataataataatttttcagcaa	6300	FT	repeat_unit	6062..6138
QY	6241	ggagcgaagcccgctcgtgtaaacatttttaaatgataataataatttttcagcaa	6300	FT	repeat_unit	6298..6537
Db	6301	gtgtggagaagcgcgagcgaagctgacattctgattcgatacgcgaacagctcaa	6360	FT	repeat_unit	6538..7012
QY	6301	gtgtggagaagcgcgagcgaagctgacattctgattcgatacgcgaacagctcaa	6360	FT	repeat_unit	7012..7075
Db	6361	tatgtttcgtggagaacacagtcgtcgtggatcatggttca	6400	FT	repeat_unit	7076..7153
QY	6361	tatgtttcgtggagaacacagtcgtcgtggatcatggttca	6400	FT	repeat_unit	7154..7231
RESULT	2					
ID	Q64742	standard; DNA; 7653 BP.				
AC	Q64742					
DE	23-JUN-1994	(first entry)				
KW	ced-3 (C6434T) gene.					
KW	C. elegans; ced-4; ced-3; mutant; transcriptional regulation;					
KW	embryogenesis; cell death; hydrophilic; transmembrane; region;					
KW	hydrophobic; mutation; amino acid; substitution; RNA splicing;					
KW	protein synthesis; null phenotype; calcium-binding domain; ss.					
OS	Caenorhabditis elegans.					
FH	Key	Location/Qualifiers				
FT	repeat_unit	1356..1472				
FT		/*tag= a				
FT		/rpt_type= INVERTED				
FT		/note= "Inverted w.r.t. repeat at 1490-1614"				
FT	repeat_unit	1490..1614				
FT		/*tag= b				
FT		/rpt_type= INVERTED				
FT		/note= "Inverted w.r.t. repeat at 1356-1472"				
FT	exon	2232..2366				
FT		/*tag= c				
FT		/number= Exon_1				
FT	intron	2367..2429				
FT		/*tag= d				
FT		/number= Intron_1				
FT	exon	2450..2575				
FT		/*tag= e				
FT		/number= Exon_2				
FT	intron	2576..2853				
FT		/*tag= f				
FT		/number= Intron_2				


```

QY 6061 AAACCAATCAGCATCGTCGATCTCCGCCCACTTCATCGGATGGTTGAAAGTGGCGGA 6120
Db 6121 gtgaattgctgattggtcgcagtttccagtttagaggggaatttaaaatcgccttttcca 6180
QY 6121 GTGAATTGCTGATGGTCGAGTTTTCAGTTTAGAGGGAATTTAAAAATCGCCTTTTCGA 6180
Db 6181 aaattaaaattgattttttcaatttttcgaaaaatattccgattattttatattcttt 6240
QY 6181 AAATTAAAATTGATTTTTCATTTTTCGAAAAATATTCCGATTATTTTATATTCTTT 6240
Db 6241 ggagcgaaagcccgctctgtaaacatttttaaatgataataataatttttgcagcaa 6300
QY 6241 GGAGCGAAAGCCCGCTCTGTAAACATTTTAAATGATAATAATAATTTTTCAGCAA 6300
Db 6301 gtgtggagaagaagccgagccagctgctgctgctgctgctgctgctgctgctgctgct 6360
QY 6301 GTGTGGAGAAAGAGCCGAGCCAGCTGACATTTCTGATTCGATACGCAACGACAGCTCAA 6360
Db 6361 tatgtttcgtggagaagaacagctgctgctgctgctgctgctgctgctgctgctgct 6400
QY 6361 TATGTTTCGTGGAGAAACAGTCTGCTGGATCATGGTTCA 6400

```

RESULT 4

```

ID Q64743 standard; DNA; 7653 BP.
AC Q64743;
DT 23-JUN-1994 (first entry)
DE ced-3 (C6485F) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key
FT Location/Qualifiers
FT repeat_unit
FT /*tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT repeat_unit
FT /*tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT exon
FT /*tag= c
FT /number= Exon_1
FT 2367..2429
FT /*tag= d
FT /number= Intron_1
FT 2450..2575
FT /*tag= e
FT /number= Exon_2
FT 2576..2853
FT /*tag= f
FT /number= Intron_2
FT 2854..3107
FT /*tag= g
FT /number= Exon_3
FT 3108..4302
FT /*tag= h
FT /number= Intron_3
FT 3126..3243
FT /*tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit
FT /*tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT repeat_unit
FT /*tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"

```

```

FT repeat_unit
FT /*tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT exon
FT /*tag= m
FT /number= Exon_4
FT 4635..5546
FT /*tag= n
FT /number= Intron_4
FT 4688..4719
FT /*tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT repeat_unit
FT /*tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon
FT /*tag= q
FT /number= Exon_5
FT 5761..5814
FT /*tag= r
FT /number= Intron_5
FT 5815..5942
FT /*tag= s
FT /number= Exon_6
FT 5943..6297
FT /*tag= t
FT /number= Intron_6
FT 6062..6138
FT /*tag= u
FT /rpt_type= INVERTED
FT 6298..6537
FT /*tag= v
FT /number= Exon_7
FT 6485
FT /*tag= w
FT /note= "C>T, from allele n2430"
FT 6538..7012
FT /*tag= x
FT /number= Intron_7
FT 6567..6625
FT /*tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit
FT /*tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon
FT /*tag= aa
FT /number= Exon_8
FT W09325685-A.
FT 23-DEC-1993.
FT 14-JUN-1993; U05701.
FT 12-JUN-1992; US-897788.
FT 20-NOV-1992; US-979638.
FT (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT Horvitz HR, Shahan S, Yuan J;
FT WPI; 94-007542/01.
FT P-PSDB; R53286.
FT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT develop agents to increase or prevent cell death in organisms
FT Claim 14; Fig 4; 127pp; English.
FT The sequences given in Q64735-45 represent mutations of the C. elegans
FT ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT was most abundant in embryos, but was also detected in larvae and young
FT adults, suggesting that ced-3 is expressed not only in cells undergoing
FT cell death. The four largest introns as well as sequences 5' of the
FT start codon contain repetitive elements, some of which have been
FT characterised in non-coding regions of other C. elegans genes, such
FT as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in

```


Db 6361 tatgttcgtgagaaacagtctcgtgatcatggttca 6400
QY |||||
6361 TATGTTTCGTGAGAAACAGTCTCGTGATCATGTTCA 6400

RESULT 6
ID Q54666 standard; DNA; 7653 BP.
AC Q54666;
DT 23-JUN-1994 (first entry)
DE ced-3 gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /*tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT 1490..1614
FT /*tag= b
FT /rpt_type= INVERTED
FT /note= "inverted w.r.t. repeat at 1356-1472"
FT 2232..2366
FT /*tag= c
FT /number= Exon_1
FT 2310
FT /*tag= d
FT /note= "C>T, from allele n1040"
FT 2367..2429
FT /*tag= e
FT /number= Intron_1
FT 2430..2575
FT /*tag= f
FT /number= Exon_2
FT 2487
FT /*tag= g
FT /note= "G>A, fromm allele n718"
FT 2576..2853
FT /*tag= h
FT /number= Intron_2
FT 2854..3107
FT /*tag= i
FT /number= Exon_3
FT 3108..4302
FT /*tag= j
FT /number= Intron_3
FT 3126..3243
FT /*tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT 3329..3396
FT /*tag= l
FT /rpt_type= INVERTED
FT /note= "inverted w.r.t. repeat at 3126-3243"
FT 3487..3759
FT /*tag= m
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT 3782..4070
FT /*tag= n
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT 4303..4634
FT /*tag= o
FT /number= Exon_4
FT 4635..5546
FT /*tag= p
FT /number= Intron_4
FT 4688..4719
FT /*tag= q

FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5221..5330
FT /*tag= r
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5547..5760
FT /*tag= s
FT /number= Exon_5
FT 5757
FT /*tag= t
FT /note= "G>A, from allele n2433"
FT 5761..5814
FT /*tag= u
FT /number= Intron_5
FT 5815..5942
FT /*tag= v
FT /number= Exon_6
FT 5940
FT /*tag= w
FT /note= "C>T, from allele n1165"
FT 5943..6297
FT /*tag= x
FT /number= Intron_6
FT 6062..6138
FT /*tag= y
FT /rpt_type= INVERTED
FT 6297
FT /*tag= z
FT /note= "G>A, fromm allele n717"
FT 6298..6537
FT /*tag= aa
FT /number= Exon_7
FT 6322
FT /*tag= ab
FT /note= "C>T, from allele n1949"
FT 6372
FT /*tag= ac
FT /note= "G>A, from allele n1286"
FT 6434
FT /*tag= ad
FT /note= "C>T, fromm alleles n1129 and n1164"
FT 6485
FT /*tag= ae
FT /note= "C>T, from allele n2430"
FT 6535
FT /*tag= af
FT /note= "G>A, from allele n2426"
FT 6538..7012
FT /*tag= ag
FT /number= Intron_7
FT 6567..6625
FT /*tag= ah
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT 6905..6965
FT /*tag= ai
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT 7013..7075
FT /*tag= aj
FT /number= Exon_8
FT 7020
FT /*tag= ak
FT /note= "C>T, from allele n1163"
FT WO9325685-A.
PN 23-DEC-1993.
PD 14-JUN-1993; U05701.
PE 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PR (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.


```

Db 6181 aaattaaaaattgatttttttcaatttttttcgaaaaaatattccgattattttatattcttt 6240
    |||||||
QY 6181 AAATTAATAATGATTTTTCATTTTTCGAAAAATATCCGATTATTTTATATCTTT 6240
    |||||||
Db 6241 ggagcgaagcccgctcctgtaaacatttttaaatgataataataaaatttttgcagcaa 6300
    |||||||
QY 6241 GGAGCGAAAGCCCGCTCTGTAACATTTTAAATGATAATTAATAAATTTTTCAGCAA 6300
    |||||||
Db 6301 gtgtggagaagaagccgagcgaagctgacattctgattcgcatacgcaacagacagctcaa 6360
    |||||||
QY 6301 GTGTGGAGAAAGAGCCGAGCGCAAGCTGACATTCTGATTCGATACGCACACGACAGCTCAA 6360
    |||||||
Db 6361 tatgtttcgtggagaacacagtgctcgtggatcatggttca 6400
    |||||||
QY 6361 TATGTTTCGTGGAGAAACAGTGTCTCGTGGATCATGTTCA 6400
    |||||||

RESULT 8
ID Q54401 standard; DNA; 7653 BP.
AC Q54401;
NT 01-JUL-1994 (first entry)
DE ced-3 gene.
KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
KW interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
KW inflammatory response; nematode; diagnosis; myocardial infarction;
KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
KW pathogenic infection; hair loss; cancer; autoreactive antibody; ss.
US Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /*tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /*tag= b
FT /rpt_type= INVERTED
FT /note= "inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT /*tag= c
FT /number= Exon_1
FT mutation 2310
FT /*tag= d
FT /note= "C>T, from allele n1040"
FT intron 2367..2429
FT /*tag= e
FT /number= Intron_1
FT exon 2430..2575
FT /*tag= f
FT /number= Exon_2
FT mutation 2487
FT /*tag= g
FT /note= "G>A, fromm allele n718"
FT intron 2576..2853
FT /*tag= h
FT /number= Intron_2
FT exon 2854..3107
FT /*tag= i
FT /number= Exon_3
FT intron 3108..4302
FT /*tag= j
FT /number= Intron_3
FT repeat_unit 3126..3243
FT /*tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit 3329..3396
FT /*tag= l
FT /rpt_type= INVERTED
FT /note= "inverted w.r.t. repeat at 3126-3243"
FT repeat_unit 3487..3759
FT /*tag= m
FT /rpt_type= INVERTED

```

```

FT repeat_unit /note= "Inverted w.r.t. repeat at 3782-4070"
FT /*tag= n
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT exon 4303..4634
FT /*tag= o
FT /number= Exon_4
FT intron 4635..5546
FT /*tag= p
FT /number= Intron_4
FT repeat_unit 4688..4719
FT /*tag= q
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT repeat_unit 5221..5330
FT /*tag= r
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon 5547..5760
FT /*tag= s
FT /number= Exon_5
FT mutation 5757
FT /*tag= t
FT /note= "G>A, from allele n2433"
FT intron 5761..5814
FT /*tag= u
FT /number= Intron_5
FT exon 5815..5942
FT /*tag= v
FT /number= Exon_6
FT mutation 5940
FT /*tag= w
FT /note= "C>T, from allele n1165"
FT intron 5943..6297
FT /*tag= x
FT /number= Intron_6
FT repeat_region 6062..6138
FT /*tag= y
FT /rpt_type= INVERTED
FT mutation 6297
FT /*tag= z
FT /note= "G>A, fromm allele n717"
FT exon 6298..6537
FT /*tag= aa
FT /number= Exon_7
FT mutation 6322
FT /*tag= ab
FT /note= "C>T, from allele n1949"
FT mutation 6372
FT /*tag= ac
FT /note= "G>A, from allele n1286"
FT mutation 6434
FT /*tag= ad
FT /note= "C>T, fromm alleles n1129 and n1164"
FT mutation 6485
FT /*tag= ae
FT /note= "C>T, from allele n2430"
FT mutation 6535
FT /*tag= af
FT /note= "G>A, from allele n2426"
FT intron 6538..7012
FT /*tag= ag
FT /number= Intron_7
FT repeat_unit 6567..6625
FT /*tag= ah
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /*tag= ai
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7013..7075

```


Db 5881 gatggacaatcgagacggccattgttcaattttcttggatgtgtcgccgcaagttc 5940
|||||
QY 5881 GATGGACAATCGAGACGGCCATTGTTCATATTTCTTGGATGTGTGGCCGCAAGTTC 5940
Db 5941 aggttgcaatttaatttcttgatgagaatattccttcaaaaaatctaaatagattttt 6000
|||||
QY 5941 AGGTTGCAATTTAATTTCTTGAATGAGAAATATCCTTCAAAAAATCTAAATAGATTTT 6000
Db 6001 attccagaaagtcctcgatcgaaaaattcgatataataattacgaaatttgtgataaaatgac 6060
|||||
QY 6001 ATCCAGAAAGTCCCGATCGATCTCCGCCCACTTCATCGGATTGGTTTGAAAGTGGCGGA 6060
Db 6061 aaaccaatcagcatcgatcgatctcgccacttcctcgatcgatgggtttgaaagtgccgga 6120
|||||
QY 6061 AAACCAATCAGCATCGTCGATCTCCGCCCACTTCATCGGATTGGTTTGAAAGTGGCGGA 6120
Db 6121 gtgaattgctgattggtcgagttttcagttttcagtttagagggaatttaaaaaatcgcccttttoga 6180
|||||
QY 6121 GTGAATTGCTGATTGTCGAGTTTTCAGTTTTCGAAATAATTCGGAATTTAATAATTCCTTTTGA 6180
uu 6181 aaattaaaaattgatttttttcaattttttcagttttcagtttagagggaatttaaaaaatcgcccttttoga 6240
|||||
QY 6181 AAATTAAAAATTGATTTTTCAGTTTTCGAAATAATTCGGAATTTAATAATTCCTTTTGA 6240
Db 6241 ggagcgaaagcccgctcctgtaaacatttttaaatgataataataaatttttgcagcaa 6300
|||||
QY 6241 GGAGCGAAAGCCCGCTCCTGTAACATTTTAAATGATAATTAATAATAATTTTGCAGCAA 6300
Db 6301 gtgtggagaagaagccgagccagctgacattctgattcgatacgcaacgacagctcaa 6360
|||||
QY 6301 GTGTGGAGAAGAAGCCGAGCCAGCTGACATTTCTGATTGATACGCAACGACAGCTCAA 6360
Db 6361 tatgtttcgtggagaacacagtgctgctggatcatggttca 6400
|||||
QY 6361 TATGTTTCGTGAGAAACAGTGCTGCTGATCATGTTCA 6400

RESULT 10
ID Q64740 standard; DNA; 7653 BP.
AC Q64740;
DT 23-JUN-1994 (first entry)
DE ced-3 (C6322T) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT /tag= c
FT /number= Exon_1
FT /tag= d
FT /number= Intron_1
FT /tag= e
FT /number= Exon_2
FT /tag= f
FT /number= Intron_2
FT /tag= g
FT /number= Exon_3
FT /tag= h

FT repeat_unit
FT /number= Intron_3
FT 3126..3243
FT /tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit 3329..3396
FT /tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT repeat_unit 3487..3759
FT /tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT repeat_unit 4070..4634
FT /tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT exon 4303..4634
FT /tag= m
FT /number= Exon_4
FT 4635..5546
FT /tag= n
FT /number= Intron_4
FT 4688..4719
FT /tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon 5547..5760
FT /tag= q
FT /number= Exon_5
FT 5761..5814
FT /tag= r
FT /number= Intron_5
FT 5815..5942
FT /tag= s
FT /number= Exon_6
FT 5943..6297
FT /tag= t
FT /number= Intron_6
FT 6062..6138
FT /tag= u
FT /rpt_type= INVERTED
FT exon 6298..6537
FT /tag= v
FT /number= Exon_7
FT 6322
FT /tag= w
FT /note= "C>T, from allele n1949"
FT intron 6538..7012
FT /tag= x
FT /number= Intron_7
FT 6567..6625
FT /tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7012..7075
FT /tag= aa
FT /number= Exon_8
FT WO9325685-A.
FT 23-DEC-1993.
FT 14-JUN-1993; U05701.
FT 12-JUN-1992; US-897788.
FT 20-NOV-1992; US-979638.
FT (MASI) MASSACHUSETTS INST TECHNOLOGY.
FT Horvitz HR, Shahan S, Yuan J; PI


```
QY 6121 GTGAATTGCTGATGCTGCGCAGTTTTCAGTTTAGAGGGAATTTAAAAATCGCCTTTTCGA 6180
Db 6181 aaattaaaaattgatttttttcaatttttttgcgaaaaatattccgattatttttatattcttt 6240
QY 6181 AAATTAAAAATTGATTTTTCATTTTTCGAAAAATATTCCGATTTATTTATATTCTTT 6240
Db 6241 ggagcgaagcccgctcctgtataacatttttaaatgataataataatttttgcagcaa 6300
QY 6241 GGAGCGAAAGCCCGCTCTGTAAACATTTTAAATGATAATTAATAATAATTTTTCAGCAA 6300
Db 6301 gtgtggagaaagccgagccagctgacattctgattcgatacgaacgacagctcaa 6360
QY 6301 GTGTGGAGAAAGAGCCGAGCCAGCTGACATTTCTGATTTCGATACGCAACGACAGCTCAA 6360
Db 6361 tatgtttcgtgaagaacagtgctcgtggatcatggttca 6400
QY 6361 TATGTTTCGTGGAGAAACAGTGCTCGTGGATCATGTTCA 6400

RESULT 12
ID Q64739 standard; DNA: 7653 BP.
AC Q64739;
DT 23-JUN-1994 (first entry)
DE ced-3 (G6297A) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT /tag= c
FT /number= Exon_1
FT intron 2367..2429
FT /tag= d
FT /number= Intron_1
FT exon 2450..2575
FT /tag= e
FT /number= Exon_2
FT intron 2576..2853
FT /tag= f
FT /number= Intron_2
FT exon 2854..3107
FT /tag= g
FT /number= Exon_3
FT intron 3108..4302
FT /tag= h
FT /number= Intron_3
FT repeat_unit 3126..3243
FT /tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit 3329..3396
FT /tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT repeat_unit 3487..3759
FT /tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT repeat_unit 3782..4070
FT /tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT exon 4303..4634
```

```
FT /tag= m
FT /number= Exon_4
FT intron 4635..5546
FT /tag= n
FT /number= Intron_4
FT repeat_unit 4688..4719
FT /tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT repeat_unit 5221..5330
FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon 5547..5760
FT /tag= q
FT /number= Exon_5
FT intron 5761..5814
FT /tag= r
FT /number= Intron_5
FT exon 5815..5942
FT /tag= s
FT /number= Exon_6
FT intron 5943..6297
FT /tag= t
FT /number= Intron_6
FT mutation 5757
FT /tag= u
FT /note= "C>T, from allele n717"
FT repeat_region 6062..6138
FT /tag= v
FT /rpt_type= INVERTED
FT exon 6298..6537
FT /tag= w
FT /number= Exon_7
FT intron 6538..7012
FT /tag= x
FT /number= Intron_7
FT repeat_unit 6567..6625
FT /tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7012..7075
FT /tag= aa
FT /number= Exon_8
FT WO9325685-A.
FT 23-DEC-1993.
FT 14-JUN-1993; U05701.
FT 12-JUN-1992; US-897788.
FT 20-NOV-1992; US-979638.
FT (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT Horvitz HR, Shaham S, Yuan J;
FT WPI; 94-007542/01.
FT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT develop agents to increase or prevent cell death in organisms
FT Claim 14; Fig 4; 127pp; English.
FT The sequences given in Q64735-45 represent mutations of the C. elegans
FT ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT was most abundant in embryos, but was also detected in larvae and young
FT adults, suggesting that ced-3 is expressed not only in cells undergoing
FT cell death. The four largest introns as well as sequences 5' of the
FT start codon contain repetitive elements, some of which have been
FT characterised in non-coding regions of other C. elegans genes, such
FT as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
FT length. Ced-3 is highly hydrophilic with no significant hydrophobic
FT region that might be a transmembrane region. One region of Ced-3 is
FT very rich in serine. It is thought that this region is involved in
FT protein-protein interactions, similar to acid blobs in transcription
FT factors. Of the mutations which occur within the ced-3 gene, eight of
FT the mutations are missense mutations, two are nonsense mutations and
```



```
FT      /*tag= p
FT      /rpt_type= INVERTED
FT      /note= "Repeat 3"
FT      5547..5760
FT      /*tag= q
FT      /number= Exon_5
FT      5761..5814
FT      /*tag= r
FT      /number= Intron_5
FT      5815..5942
FT      /*tag= s
FT      /number= Exon_6
FT      5943..6297
FT      /*tag= t
FT      /note= "C>T, from allele n1165"
FT      5943..6297
FT      /*tag= u
FT      /number= Intron_6
FT      6062..6138
FT      /*tag= v
FT      /rpt_type= INVERTED
FT      6298..6537
FT      /*tag= w
FT      /number= Exon_7
FT      6538..7012
FT      /*tag= x
FT      /number= Intron_7
FT      6567..6625
FT      /*tag= y
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT      6905..6965
FT      /*tag= z
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      7012..7075
FT      /*tag= aa
FT      /number= Exon_8
FT      7076..7140
FT      /number= Exon_9
FT      7141..7205
FT      /number= Exon_10
FT      7206..7270
FT      /number= Exon_11
FT      7271..7335
FT      /number= Exon_12
FT      7336..7400
FT      /number= Exon_13
FT      7401..7465
FT      /number= Exon_14
FT      7466..7530
FT      /number= Exon_15
FT      7531..7595
FT      /number= Exon_16
FT      7596..7660
FT      /number= Exon_17
FT      7661..7725
FT      /number= Exon_18
FT      7726..7790
FT      /number= Exon_19
FT      7791..7855
FT      /number= Exon_20
FT      7856..7920
FT      /number= Exon_21
FT      7921..7985
FT      /number= Exon_22
FT      7986..8050
FT      /number= Exon_23
FT      8051..8115
FT      /number= Exon_24
FT      8116..8180
FT      /number= Exon_25
FT      8181..8245
FT      /number= Exon_26
FT      8246..8310
FT      /number= Exon_27
FT      8311..8375
FT      /number= Exon_28
FT      8376..8440
FT      /number= Exon_29
FT      8441..8505
FT      /number= Exon_30
FT      8506..8570
FT      /number= Exon_31
FT      8571..8635
FT      /number= Exon_32
FT      8636..8700
FT      /number= Exon_33
FT      8701..8765
FT      /number= Exon_34
FT      8766..8830
FT      /number= Exon_35
FT      8831..8895
FT      /number= Exon_36
FT      8896..8960
FT      /number= Exon_37
FT      8961..9025
FT      /number= Exon_38
FT      9026..9090
FT      /number= Exon_39
FT      9091..9155
FT      /number= Exon_40
FT      9156..9220
FT      /number= Exon_41
FT      9221..9285
FT      /number= Exon_42
FT      9286..9350
FT      /number= Exon_43
FT      9351..9415
FT      /number= Exon_44
FT      9416..9480
FT      /number= Exon_45
FT      9481..9545
FT      /number= Exon_46
FT      9546..9610
FT      /number= Exon_47
FT      9611..9675
FT      /number= Exon_48
FT      9676..9740
FT      /number= Exon_49
FT      9741..9805
FT      /number= Exon_50
FT      9806..9870
FT      /number= Exon_51
FT      9871..9935
FT      /number= Exon_52
FT      9936..10000
FT      /number= Exon_53
FT      10001..10065
FT      /number= Exon_54
FT      10066..10130
FT      /number= Exon_55
FT      10131..10195
FT      /number= Exon_56
FT      10196..10260
FT      /number= Exon_57
FT      10261..10325
FT      /number= Exon_58
FT      10326..10390
FT      /number= Exon_59
FT      10391..10455
FT      /number= Exon_60
FT      10456..10520
FT      /number= Exon_61
FT      10521..10585
FT      /number= Exon_62
FT      10586..10650
FT      /number= Exon_63
FT      10651..10715
FT      /number= Exon_64
FT      10716..10780
FT      /number= Exon_65
FT      10781..10845
FT      /number= Exon_66
FT      10846..10910
FT      /number= Exon_67
FT      10911..10975
FT      /number= Exon_68
FT      10976..11040
FT      /number= Exon_69
FT      11041..11105
FT      /number= Exon_70
FT      11106..11170
FT      /number= Exon_71
FT      11171..11235
FT      /number= Exon_72
FT      11236..11300
FT      /number= Exon_73
FT      11301..11365
FT      /number= Exon_74
FT      11366..11430
FT      /number= Exon_75
FT      11431..11495
FT      /number= Exon_76
FT      11496..11560
FT      /number= Exon_77
FT      11561..11625
FT      /number= Exon_78
FT      11626..11690
FT      /number= Exon_79
FT      11691..11755
FT      /number= Exon_80
FT      11756..11820
FT      /number= Exon_81
FT      11821..11885
FT      /number= Exon_82
FT      11886..11950
FT      /number= Exon_83
FT      11951..12015
FT      /number= Exon_84
FT      12016..12080
FT      /number= Exon_85
FT      12081..12145
FT      /number= Exon_86
FT      12146..12210
FT      /number= Exon_87
FT      12211..12275
FT      /number= Exon_88
FT      12276..12340
FT      /number= Exon_89
FT      12341..12405
FT      /number= Exon_90
FT      12406..12470
FT      /number= Exon_91
FT      12471..12535
FT      /number= Exon_92
FT      12536..12600
FT      /number= Exon_93
FT      12601..12665
FT      /number= Exon_94
FT      12666..12730
FT      /number= Exon_95
FT      12731..12795
FT      /number= Exon_96
FT      12796..12860
FT      /number= Exon_97
FT      12861..12925
FT      /number= Exon_98
FT      12926..12990
FT      /number= Exon_99
FT      12991..13055
FT      /number= Exon_100
FT      13056..13120
FT      /number= Exon_101
FT      13121..13185
FT      /number= Exon_102
FT      13186..13250
FT      /number= Exon_103
FT      13251..13315
FT      /number= Exon_104
FT      13316..13380
FT      /number= Exon_105
FT      13381..13445
FT      /number= Exon_106
FT      13446..13510
FT      /number= Exon_107
FT      13511..13575
FT      /number= Exon_108
FT      13576..13640
FT      /number= Exon_109
FT      13641..13705
FT      /number= Exon_110
FT      13706..13770
FT      /number= Exon_111
FT      13771..13835
FT      /number= Exon_112
FT      13836..13900
FT      /number= Exon_113
FT      13901..13965
FT      /number= Exon_114
FT      13966..14030
FT      /number= Exon_115
FT      14031..14095
FT      /number= Exon_116
FT      14096..14160
FT      /number= Exon_117
FT      14161..14225
FT      /number= Exon_118
FT      14226..14290
FT      /number= Exon_119
FT      14291..14355
FT      /number= Exon_120
FT      14356..14420
FT      /number= Exon_121
FT      14421..14485
FT      /number= Exon_122
FT      14486..14550
FT      /number= Exon_123
FT      14551..14615
FT      /number= Exon_124
FT      14616..14680
FT      /number= Exon_125
FT      14681..14745
FT      /number= Exon_126
FT      14746..14810
FT      /number= Exon_127
FT      14811..14875
FT      /number= Exon_128
FT      14876..14940
FT      /number= Exon_129
FT      14941..15005
FT      /number= Exon_130
FT      15006..15070
FT      /number= Exon_131
FT      15071..15135
FT      /number= Exon_132
FT      15136..15200
FT      /number= Exon_133
FT      15201..15265
FT      /number= Exon_134
FT      15266..15330
FT      /number= Exon_135
FT      15331..15395
FT      /number= Exon_136
FT      15396..15460
FT      /number= Exon_137
FT      15461..15525
FT      /number= Exon_138
FT      15526..15590
FT      /number= Exon_139
FT      15591..15655
FT      /number= Exon_140
FT      15656..15720
FT      /number= Exon_141
FT      15721..15785
FT      /number= Exon_142
FT      15786..15850
FT      /number= Exon_143
FT      15851..15915
FT      /number= Exon_144
FT      15916..15980
FT      /number= Exon_145
FT      15981..16045
FT      /number= Exon_146
FT      16046..16110
FT      /number= Exon_147
FT      16111..16175
FT      /number= Exon_148
FT      16176..16240
FT      /number= Exon_149
FT      16241..16305
FT      /number= Exon_150
FT      16306..16370
FT      /number= Exon_151
FT      16371..16435
FT      /number= Exon_152
FT      16436..16500
FT      /number= Exon_153
FT      16501..16565
FT      /number= Exon_154
FT      16566..16630
FT      /number= Exon_155
FT      16631..16695
FT      /number= Exon_156
FT      16696..16760
FT      /number= Exon_157
FT      16761..16825
FT      /number= Exon_158
FT      16826..16890
FT      /number= Exon_159
FT      16891..16955
FT      /number= Exon_160
FT      16956..17020
FT      /number= Exon_161
FT      17021..17085
FT      /number= Exon_162
FT      17086..17150
FT      /number= Exon_163
FT      17151..17215
FT      /number= Exon_164
FT      17216..17280
FT      /number= Exon_165
FT      17281..17345
FT      /number= Exon_166
FT      17346..17410
FT      /number= Exon_167
FT      17411..17475
FT      /number= Exon_168
FT      17476..17540
FT      /number= Exon_169
FT      17541..17605
FT      /number= Exon_170
FT      17606..17670
FT      /number= Exon_171
FT      17671..17735
FT      /number= Exon_172
FT      17736..17800
FT      /number= Exon_173
FT      17801..17865
FT      /number= Exon_174
FT      17866..17930
FT      /number= Exon_175
FT      17931..17995
FT      /number= Exon_176
FT      17996..18060
FT      /number= Exon_177
FT      18061..18125
FT      /number= Exon_178
FT      18126..18190
FT      /number= Exon_179
FT      18191..18255
FT      /number= Exon_180
FT      18256..18320
FT      /number= Exon_181
FT      18321..18385
FT      /number= Exon_182
FT      18386..18450
FT      /number= Exon_183
FT      18451..18515
FT      /number= Exon_184
FT      18516..18580
FT      /number= Exon_185
FT      18581..18645
FT      /number= Exon_186
FT      18646..18710
FT      /number= Exon_187
FT      18711..18775
FT      /number= Exon_188
FT      18776..18840
FT      /number= Exon_189
FT      18841..18905
FT      /number= Exon_190
FT      18906..18970
FT      /number= Exon_191
FT      18971..19035
FT      /number= Exon_192
FT      19036..19100
FT      /number= Exon_193
FT      19101..19165
FT      /number= Exon_194
FT      19166..19230
FT      /number= Exon_195
FT      19231..19295
FT      /number= Exon_196
FT      19296..19360
FT      /number= Exon_197
FT      19361..19425
FT      /number= Exon_198
FT      19426..19490
FT      /number= Exon_199
FT      19491..19555
FT      /number= Exon_200
FT      19556..19620
FT      /number= Exon_201
FT      19621..19685
FT      /number= Exon_202
FT      19686..19750
FT      /number= Exon_203
FT      19751..19815
FT      /number= Exon_204
FT      19816..19880
FT      /number= Exon_205
FT      19881..19945
FT      /number= Exon_206
FT      19946..20010
FT      /number= Exon_207
FT      20011..20075
FT      /number= Exon_208
FT      20076..20140
FT      /number= Exon_209
FT      20141..20205
FT      /number= Exon_210
FT      20206..20270
FT      /number= Exon_211
FT      20271..20335
FT      /number= Exon_212
FT      20336..20400
FT      /number= Exon_213
FT      20401..20465
FT      /number= Exon_214
FT      20466..20530
FT      /number= Exon_215
FT      20531..20595
FT      /number= Exon_216
FT      20596..20660
FT      /number= Exon_217
FT      20661..20725
FT      /number= Exon_218
FT      20726..20790
FT      /number= Exon_219
FT      20791..20855
FT      /number= Exon_220
FT      20856..20920
FT      /number= Exon_221
FT      20921..20985
FT      /number= Exon_222
FT      20986..21050
FT      /number= Exon_223
FT      21051..21115
FT      /number= Exon_224
FT      21116..21180
FT      /number= Exon_225
FT      21181..21245
FT      /number= Exon_226
FT      21246..21310
FT      /number= Exon_227
FT      21311..21375
FT      /number= Exon_228
FT      21376..21440
FT      /number= Exon_229
FT      21441..21505
FT      /number= Exon_230
FT      21506..21570
FT      /number= Exon_231
FT      21571..21635
FT      /number= Exon_232
FT      21636..21700
FT      /number= Exon_233
FT      21701..21765
FT      /number= Exon_234
FT      21766..21830
FT      /number= Exon_235
FT      21831..21895
FT      /number= Exon_236
FT      21896..21960
FT      /number= Exon_237
FT      21961..22025
FT      /number= Exon_238
FT      22026..22090
FT      /number= Exon_239
FT      22091..22155
FT      /number= Exon_240
FT      22156..22220
FT      /number= Exon_241
FT      22221..22285
FT      /number= Exon_242
FT      22286..22350
FT      /number= Exon_243
FT      22351..22415
FT      /number= Exon_244
FT      22416..22480
FT      /number= Exon_245
FT      22481..22545
FT      /number= Exon_246
FT      22546..22610
FT      /number= Exon_247
FT      22611..22675
FT      /number= Exon_248
FT      22676..22740
FT      /number= Exon_249
FT      22741..22805
FT      /number= Exon_250
FT      22806..22870
FT      /number= Exon_251
FT      22871..22935
FT      /number= Exon_252
FT      22936..23000
FT      /number= Exon_253
FT      23001..23065
FT      /number= Exon_254
FT      23066..23130
FT      /number= Exon_255
FT      23131..23195
FT      /number= Exon_256
FT      23196..23260
FT      /number= Exon_257
FT      23261..23325
FT      /number= Exon_258
FT      23326..23390
FT      /number= Exon_259
FT      23391..23455
FT      /number= Exon_260
FT      23456..23520
FT      /number= Exon_261
FT      23521..23585
FT      /number= Exon_262
FT      23586..23650
FT      /number= Exon_263
FT      23651..23715
FT      /number= Exon_264
FT      23716..23780
FT      /number= Exon_265
FT      23781..23845
FT      /number= Exon_266
FT      23846..23910
FT      /number= Exon_267
FT      23911..23975
FT      /number= Exon_268
FT      23976..24040
FT      /number= Exon_269
FT      24041..24105
FT      /number= Exon_270
FT      24106..24170
FT      /number= Exon_271
FT      24171..24235
FT      /number= Exon_272
FT      24236..24300
FT      /number= Exon_273
FT      24301..24365
FT      /number= Exon_274
FT      24366..24430
FT      /number= Exon_275
FT      24431..24495
FT      /number= Exon_276
FT      24496..24560
FT      /number= Exon_277
FT      24561..24625
FT      /number= Exon_278
FT      24626..24690
FT      /number= Exon_279
FT      24691..24755
FT      /number= Exon_280
FT      24756..24820
FT      /number= Exon_281
FT      24821..24885
FT      /number= Exon_282
FT      24886..24950
FT      /number= Exon_283
FT      24951..25015
FT      /number= Exon_284
FT      25016..25080
FT      /number= Exon_285
FT      25081..25145
FT      /number= Exon_286
FT      25146..25210
FT      /number= Exon_287
FT      25211..25275
FT      /number= Exon_288
FT      25276..25340
FT      /number= Exon_289
FT      25341..25405
FT      /number= Exon_290
FT      25406..25470
FT      /number= Exon_291
FT      25471..25535
FT      /number= Exon_292
FT      25536..25600
FT      /number= Exon_293
FT      25601..25665
FT      /number= Exon_294
FT      25666..25730
FT      /number= Exon_295
FT      25731..25795
FT      /number= Exon_296
FT      25796..25860
FT      /number= Exon_297
FT      25861..25925
FT      /number= Exon_298
FT      25926..25990
FT      /number= Exon_299
FT      25991..26055
FT      /number= Exon_300
FT      26056..26120
FT      /number= Exon_301
FT      26121..26185
FT      /number= Exon_302
FT      26186..26250
FT      /number= Exon_303
FT      26251..26315
FT      /number= Exon_304
FT      26316..26380
FT      /number= Exon_305
FT      26381..26445
FT      /number= Exon_306
FT      26446..26510
FT      /number= Exon_307
FT      26511..26575
FT      /number= Exon_308
FT      26576..26640
FT      /number= Exon_309
FT      26641..26705
FT      /number= Exon_310
FT      26706..26770
FT      /number= Exon_311
FT      26771..26835
FT      /number= Exon_312
FT      26836..26900
FT      /number= Exon_313
FT      26901..26965
FT      /number= Exon_314
FT      26966..27030
FT      /number= Exon_315
FT      27031..27095
FT      /number= Exon_316
FT      27096..27160
FT      /number= Exon_317
FT      27161..27225
FT      /number= Exon_318
FT      27226..27290
FT      /number= Exon_319
FT      27291..27355
FT      /number= Exon_320
FT      27356..27420
FT      /number= Exon_321
FT      27421..27485
FT      /number= Exon_322
FT      27486..27550
FT      /number= Exon_323
FT      27551..27615
FT      /number= Exon_324
FT      27616..27680
FT      /number= Exon_325
FT      27681..27745
FT      /number= Exon_326
FT      27746..27810
FT      /number= Exon_327
FT      27811..27875
FT      /number= Exon_328
FT      27876..27940
FT      /number= Exon_329
FT      27941..28005
FT      /number= Exon_330
FT      28006..28070
FT      /number= Exon_331
FT      28071..28135
FT      /number= Exon_332
FT      28136..28200
FT      /number= Exon_333
FT      28201..28265
FT      /number= Exon_334
FT      28266..28330
FT      /number= Exon_335
FT      28331..28395
FT      /number= Exon_336
FT      28396..28460
FT      /number= Exon_337
FT      28461..28525
FT      /number= Exon_338
FT      28526..28590
FT      /number= Exon_339
FT      28591..28655
FT      /number= Exon_340
FT      28656..28720
FT      /number= Exon_341
FT      28721..28785
FT      /number= Exon_342
FT      28786..28850
FT      /number= Exon_343
FT      28851..28915
FT      /number= Exon_344
FT      28916..28980
FT      /number= Exon_345
FT      28981..29045
FT      /number= Exon_346
FT      29046..29110
FT      /number= Exon_347
FT      29111..29175
FT      /number= Exon_348
FT      29176..29240
FT      /number= Exon_349
FT      29241..29305
FT      /number= Exon_350
FT      29306..29370
FT      /number= Exon_351
FT      29371..29435
FT      /number= Exon_352
FT      29436..29500
FT      /number= Exon_353
FT      29501..29565
FT      /number= Exon_354
FT      29566..29630
FT      /number= Exon_355
FT      29631..29695
FT      /number= Exon_356
FT      29696..29760
FT      /number= Exon_357
FT      29761..29825
FT      /number= Exon_358
FT      29826..29890
FT      /number= Exon_359
FT      29891..29955
FT      /number= Exon_360
FT      29956..30020
FT      /number= Exon_361
FT      30021..30085
FT      /number= Exon_362
FT      30086..30150
FT      /number= Exon_363
FT      30151..30215
FT      /number= Exon_364
FT      30216..30280
FT      /number= Exon_365
FT      30281..30345
FT      /number= Exon_366
FT      30346..30410
FT      /number= Exon_367
FT      30411..30475
FT      /number= Exon_368
FT      30476..30540
FT      /number= Exon_369
FT      30541..30605
FT      /number= Exon_370
FT      30606..30670
FT      /number= Exon_371
FT      30671..30735
FT      /number= Exon_372
FT      30736..30800
FT      /number= Exon_373
FT      30801..30865
FT      /number= Exon_374
FT      30866..30930
FT      /number= Exon_375
FT      30931..30995
FT      /number= Exon_376
FT      30996..31060
FT      /number= Exon_377
FT      31061..31125
FT      /number= Exon_378
FT      31126..31190
FT      /number= Exon_379
FT      31191..31255
FT      /number= Exon_380
FT      31256..31320
FT      /number= Exon_381
FT      31321..31385
FT      /number= Exon_382
FT      31386..31450
FT      /number= Exon_383
FT      31451..31515
FT      /number= Exon_384
FT      31516..31580
FT      /number= Exon_385
FT      31581..31645
FT      /number= Exon_386
FT      31646..31710
FT      /number= Exon_387
FT      31711..31775
FT      /number= Exon_388
FT      31776..31840
FT      /number= Exon_389
FT      31841..31905
FT      /number= Exon_390
FT      31906..31970
FT      /number= Exon_391
FT      31971..32035
FT      /number= Exon_392
FT      32036..32100
FT      /number= Exon_393
FT      32101..32165
FT      /number= Exon_394
FT      32166..32230
FT      /number= Exon_395
FT      32231..32295
FT      /number= Exon_396
FT      32296..32360
FT      /number= Exon_397
FT      32361..32425
FT      /number= Exon_398
FT      32426..32490
FT      /number= Exon_399
FT      32491..32555
FT      /number= Exon_400
FT      32556..32620
FT      /number= Exon_401
FT      32621..32685
FT      /number= Exon_402
FT      32686..32750
FT      /number= Exon_403
FT      32751..32815
FT      /number= Exon_404
FT      32816..32880
FT      /number= Exon_405
FT      32881..32945
FT      /number= Exon_406
FT      32946..33010
FT      /number= Exon_407
FT      33011..33075
FT      /number= Exon_408
FT      33076..33140
FT      /number= Exon_409
FT      33141..33205
FT      /number= Exon_410
FT      33206..33270
FT      /number= Exon_411
FT      33271..33335
FT      /number= Exon_412
FT      33336..33400
FT      /number= Exon_413
FT      33401..33465
FT      /number= Exon_414
FT      33466..33530
FT      /number= Exon_415
FT      33531..33595
FT      /number= Exon_416
FT      33596..33660
FT      /number= Exon_417
FT      33661..33725
FT      /number= Exon_418
FT      33726..33790
FT      /number= Exon_419
FT      33791..33855
FT      /number= Exon_420
FT      33856..33920
FT      /number= Exon_421
FT      33921..33985
FT      /number= Exon_422
FT      33986..34050
FT      /number= Exon_423
FT      34051..34115
FT      /number= Exon_424
FT      34116..34180
FT      /number= Exon_425
FT      34181..34245
FT      /number= Exon_426
FT      34246..34310
FT      /number= Exon_427
FT      34311..34375
FT      /number= Exon_428
FT      34376..34440
FT      /number= Exon_429
FT      34441..34505
FT      /number= Exon_430
FT      34506..34570
FT      /number= Exon_431
FT      34571..34635
FT      /number= Exon_432
FT      34636..34700
FT      /number= Exon_433
FT      34701..34765
FT      /number= Exon_434
FT      34766..34830
FT      /number= Exon_435
FT      34831..34895
FT      /number= Exon_436
FT      34896..34960
FT      /number= Exon_437
FT      34961..35025
FT      /number= Exon_438
FT      35026..35090
FT      /number= Exon_439
FT      35091..35155
FT      /number= Exon_440
FT      35156..35220
FT      /number= Exon_441
FT      35221..35285
FT      /number= Exon_442
FT      35286..35350
FT      /number= Exon_443
FT      35351..35415
FT      /number= Exon_444
FT      35416..35480
FT      /number= Exon_445
FT      35481..35545
FT      /number= Exon_446
FT      35546..35610
FT      /number= Exon_447
FT      35611..35675
FT      /number= Exon_448
FT      35676..35740
FT      /number= Exon_449
FT      35741..35805
FT      /number= Exon_450
FT      35806..35870
FT      /number= Exon_451
FT      35871..35935
FT      /number= Exon_4
```



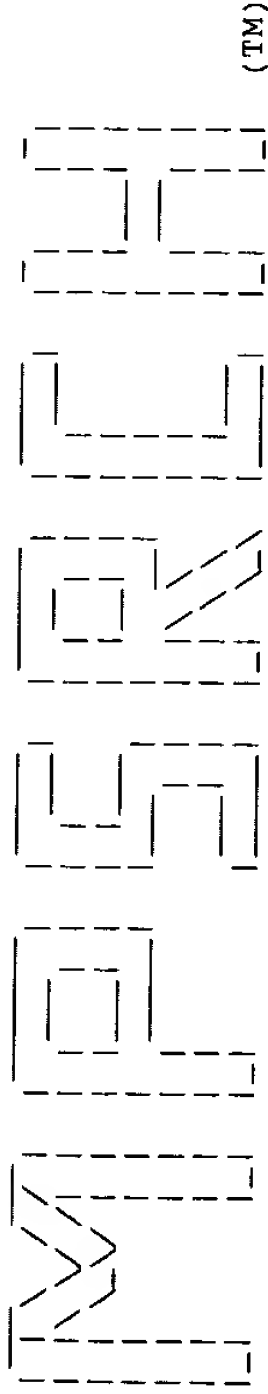
```
PT replication, maintaining circular plasmids and in assembly of human  
PT artificial chromosomes  
PS Claim 1; Page 42; 54pp; English.  
CC This sequence represents a human or mammalian DNA replication origin  
CC consensus sequences of the invention, designated uniorconsensus.  
CC Administration of the consensus sequence or an anti-gene (comprising a  
CC double stranded copy of the consensus) is used to inhibit DNA replication  
CC in vivo or in vitro. The consensus sequences can also be inserted into an  
CC expression vector, used subsequently for in vitro transfection of  
CC mammalian cells, to control initiation of DNA replication. They can also  
CC be used to maintain circular plasmids that are capable of  
CC semi-conservative replication in proliferating mammalian cells, or  
CC inserted into mammalian or human artificial chromosome vectors for gene  
CC therapy. Particularly, they are used to create shuttle vector constructs  
CC for defining the essential mammalian elements required for maintenance of  
CC chromosomal function. The consensus sequence can be combined with cloned  
CC human telomeres and large centromeric blocks for assembly of human  
CC artificial chromosomes and maintained as bacterial plasmids, circular or  
CC linear, large or small yeast artificial chromosomes (YACs) or as episomal  
CC elements.
```

```
SQ Sequence      91 BP;       15 A;        1 C;         4 G;          7 T;  
  
Query Match              4.3%; Score 43; DB 46; Length 91;  
Best Local Similarity   18.7%; Pred. No. 1.06e-07;  
Matches                17; Conservative 50; Mismatches 24; Indels 0; Gaps 0;
```

```
Ddb    1 awmtwaakrawrwwkddavwvvgakrwkwvwvrassacmdwkkaaktwggtwarrywkg 60  
      |:|:|| | : ::||:: ||::: : : : :: | ::| :  
Cp     5542 AAATTAATAATCATGTTTACATTAGGACACAGGCTGGGAATTTTTTCGGGTTCCTTTTG 5483
```

```
Ddb    61 rkmwtwkawsdatakakwwwkdakwmwrkttt 91  
      :>::|:: :>::| : :: : :: : :::  
Cp     5482 GTATTTTGCAGAAAATTGATTTTTTTTGGTTT 5452
```

```
Search completed: Sat Aug 7 21:34:12 1999  
Job time : 240 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sat Aug 7 21:00:54 1999; MasPar time 1720.83 Seconds
Tabular output not generated. 1361.646 Million cell updates/sec

Title: >US-08-287-669-18
Description: (5401-6400) from US08287669.seq (7 of 10)
Perfect Score: 1000
N.A. Sequence: 5401 ATTCAAAAAAAGTCGAAT.....TGCTCGTGATCATGTTCA 6400
Comp: TAAGTTTTTTTTCAGCTTA.....ACGAGCACCTAGTACCAAGT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

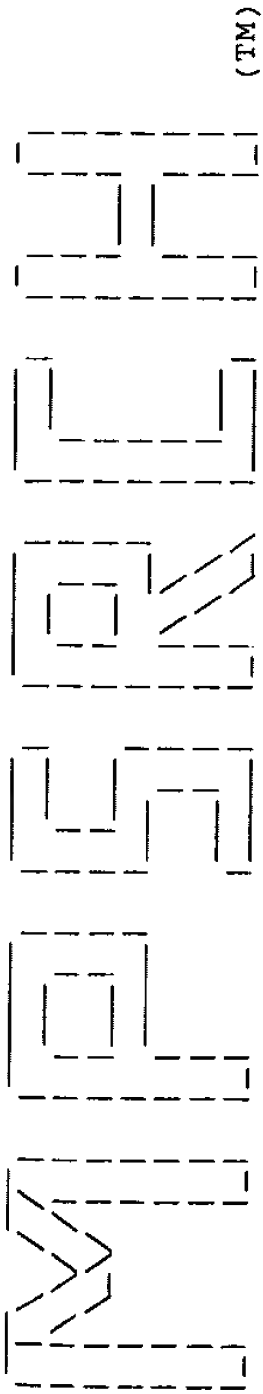
Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: embl-est58
1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
6:em_est9 7:em_gss1
Database: genbank-est111
8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
17:gb_est18 18:gb_est19 19:gb_est20 20:gb_est21
21:gb_est22 22:gb_est23 23:gb_est24 24:gb_est25
25:gb_est26 26:gb_est27 27:gb_est28 28:gb_est29
29:gb_est30 30:gb_est31 31:gb_est32 32:gb_est33 33:gb_est34
34:gb_est35 35:gb_est36 36:gb_est37 37:gb_est38 38:gb_est39
39:gb_est40 40:gb_est41 41:gb_est42 42:gb_est43 43:gb_est44
44:gb_est45 45:gb_est46 46:gb_est47 47:gb_est48 48:gb_est49
49:gb_est50 50:gb_est51 51:gb_est52 52:gb_est53 53:gb_est54
54:gb_est55 55:gb_est56 56:gb_est57 57:gb_est58 58:gb_est59
59:gb_est60 60:gb_est61 61:gb_est62 62:gb_est63 63:gb_est64
64:gb_est65 65:gb_est66 66:gb_est67 67:gb_est68 68:gb_est69
69:gb_est70 70:gb_est71 71:gb_est72 72:gb_est73 73:gb_est74
74:gb_est75 75:gb_est76 76:gb_est77 77:gb_est78 78:gb_est79
79:gb_est80 80:gb_est81 81:gb_est82 82:gb_est83 83:gb_est84
84:gb_est85 85:gb_est86 86:gb_est87 87:gb_est88 88:gb_est89
89:gb_est90 90:gb_est91 91:gb_est92 92:gb_est93 93:gb_est94
94:gb_est95 95:gb_est96 96:gb_est97 97:gb_est98 98:gb_est99
99:gb_est100 100:gb_est101 101:gb_est102 102:gb_est103
103:gb_est104 104:gb_est105 105:gb_est106 106:gb_est107
107:gb_est108 108:gb_est109 109:gb_est110 110:gb_est111
111:gb_est112 112:gb_est113 113:gb_est114 114:gb_est115
115:gb_est116 116:gb_est117 117:gb_est118 118:gb_est119
119:gb_est120 120:gb_est121 121:gb_est122 122:gb_est123
123:gb_est124 124:gb_est125 125:gb_est126 126:gb_est127
127:gb_est128 128:gb_est129 129:gb_est130 130:gb_est131
131:gb_est132 132:gb_est133 133:gb_est134 134:gb_est135
135:gb_est136 136:gb_est137 137:gb_est138 138:gb_est139
139:gb_est140 140:gb_est141 141:gb_est142 142:gb_est143
143:gb_est144 144:gb_est145 145:gb_est146 146:gb_est147
147:gb_est148 148:gb_est149 149:gb_est150 150:gb_est151
151:gb_est152 152:gb_est153 153:gb_est154 154:gb_est155
155:gb_est156 156:gb_est157 157:gb_est158 158:gb_est159
159:gb_est160 160:gb_est161 161:gb_est162 162:gb_est163
163:gb_est164 164:gb_est165 165:gb_est166 166:gb_est167
167:gb_est168 168:gb_est169 169:gb_est170 170:gb_est171
171:gb_est172 172:gb_est173 173:gb_est174 174:gb_est175
175:gb_est176 176:gb_est177 177:gb_est178 178:gb_est179
179:gb_est180 180:gb_est181 181:gb_est182 182:gb_est183
183:gb_est184 184:gb_est185 185:gb_est186 186:gb_est187
187:gb_est188 188:gb_est189 189:gb_est190 190:gb_est191
191:gb_est192 192:gb_est193 193:gb_est194 194:gb_est195
195:gb_est196 196:gb_est197 197:gb_est198 198:gb_est199
199:gb_est200 200:gb_est201 201:gb_est202 202:gb_est203
203:gb_est204 204:gb_est205 205:gb_est206 206:gb_est207
207:gb_est208 208:gb_est209 209:gb_est210 210:gb_est211
211:gb_est212 212:gb_est213 213:gb_est214 214:gb_est215
215:gb_est216 216:gb_est217 217:gb_est218 218:gb_est219
219:gb_est220 220:gb_est221 221:gb_est222 222:gb_est223
223:gb_est224 224:gb_est225 225:gb_est226 226:gb_est227
227:gb_est228 228:gb_est229 229:gb_est230 230:gb_est231
231:gb_est232 232:gb_est233 233:gb_est234 234:gb_est235
235:gb_est236 236:gb_est237 237:gb_est238 238:gb_est239
239:gb_est240 240:gb_est241 241:gb_est242 242:gb_est243
243:gb_est244 244:gb_est245 245:gb_est246 246:gb_est247
247:gb_est248 248:gb_est249 249:gb_est250 250:gb_est251
251:gb_est252 252:gb_est253 253:gb_est254 254:gb_est255
255:gb_est256 256:gb_est257 257:gb_est258 258:gb_est259
259:gb_est260 260:gb_est261 261:gb_est262 262:gb_est263
263:gb_est264 264:gb_est265 265:gb_est266 266:gb_est267
267:gb_est268 268:gb_est269 269:gb_est270 270:gb_est271
271:gb_est272 272:gb_est273 273:gb_est274 274:gb_est275
275:gb_est276 276:gb_est277 277:gb_est278 278:gb_est279
279:gb_est280 280:gb_est281 281:gb_est282 282:gb_est283
283:gb_est284 284:gb_est285 285:gb_est286 286:gb_est287
287:gb_est288 288:gb_est289 289:gb_est290 290:gb_est291
291:gb_est292 292:gb_est293 293:gb_est294 294:gb_est295
295:gb_est296 296:gb_est297 297:gb_est298 298:gb_est299
299:gb_est300 300:gb_est301 301:gb_est302 302:gb_est303
303:gb_est304 304:gb_est305 305:gb_est306 306:gb_est307
307:gb_est308 308:gb_est309 309:gb_est310 310:gb_est311
311:gb_est312 312:gb_est313 313:gb_est314 314:gb_est315
315:gb_est316 316:gb_est317 317:gb_est318 318:gb_est319
319:gb_est320 320:gb_est321 321:gb_est322 322:gb_est323
323:gb_est324 324:gb_est325 325:gb_est326 326:gb_est327
327:gb_est328 328:gb_est329 329:gb_est330 330:gb_est331
331:gb_est332 332:gb_est333 333:gb_est334 334:gb_est335
335:gb_est336 336:gb_est337 337:gb_est338 338:gb_est339
339:gb_est340 340:gb_est341 341:gb_est342 342:gb_est343
343:gb_est344 344:gb_est345 345:gb_est346 346:gb_est347
347:gb_est348 348:gb_est349 349:gb_est350 350:gb_est351
351:gb_est352 352:gb_est353 353:gb_est354 354:gb_est355
355:gb_est356 356:gb_est357 357:gb_est358 358:gb_est359
359:gb_est360 360:gb_est361 361:gb_est362 362:gb_est363
363:gb_est364 364:gb_est365 365:gb_est366 366:gb_est367
367:gb_est368 368:gb_est369 369:gb_est370 370:gb_est371
371:gb_est372 372:gb_est373 373:gb_est374 374:gb_est375
375:gb_est376 376:gb_est377 377:gb_est378 378:gb_est379
379:gb_est380 380:gb_est381 381:gb_est382 382:gb_est383
383:gb_est384 384:gb_est385 385:gb_est386 386:gb_est387
387:gb_est388 388:gb_est389 389:gb_est390 390:gb_est391
391:gb_est392 392:gb_est393 393:gb_est394 394:gb_est395
395:gb_est396 396:gb_est397 397:gb_est398 398:gb_est399
399:gb_est400 400:gb_est401 401:gb_est402 402:gb_est403
403:gb_est404 404:gb_est405 405:gb_est406 406:gb_est407
407:gb_est408 408:gb_est409 409:gb_est410 410:gb_est411
411:gb_est412 412:gb_est413 413:gb_est414 414:gb_est415
415:gb_est416 416:gb_est417 417:gb_est418 418:gb_est419
419:gb_est420 420:gb_est421 421:gb_est422 422:gb_est423
423:gb_est424 424:gb_est425 425:gb_est426 426:gb_est427
427:gb_est428 428:gb_est429 429:gb_est430 430:gb_est431
431:gb_est432 432:gb_est433 433:gb_est434 434:gb_est435
435:gb_est436 436:gb_est437 437:gb_est438 438:gb_est439
439:gb_est440 440:gb_est441 441:gb_est442 442:gb_est443
443:gb_est444 444:gb_est445 445:gb_est446 446:gb_est447
447:gb_est448 448:gb_est449 449:gb_est450 450:gb_est451
451:gb_est452 452:gb_est453 453:gb_est454 454:gb_est455
455:gb_est456 456:gb_est457 457:gb_est458 458:gb_est459
459:gb_est460 460:gb_est461 461:gb_est462 462:gb_est463
463:gb_est464 464:gb_est465 465:gb_est466 466:gb_est467
467:gb_est468 468:gb_est469 469:gb_est470 470:gb_est471
471:gb_est472 472:gb_est473 473:gb_est474 474:gb_est475
475:gb_est476 476:gb_est477 477:gb_est478 478:gb_est479
479:gb_est480 480:gb_est481 481:gb_est482 482:gb_est483
483:gb_est484 484:gb_est485 485:gb_est486 486:gb_est487
487:gb_est488 488:gb_est489 489:gb_est490 490:gb_est491
491:gb_est492 492:gb_est493 493:gb_est494 494:gb_est495
495:gb_est496 496:gb_est497 497:gb_est498 498:gb_est499
499:gb_est500 500:gb_est501 501:gb_est502 502:gb_est503
503:gb_est504 504:gb_est505 505:gb_est506 506:gb_est507
507:gb_est508 508:gb_est509 509:gb_est510 510:gb_est511
511:gb_est512 512:gb_est513 513:gb_est514 514:gb_est515
515:gb_est516 516:gb_est517 517:gb_est518 518:gb_est519
519:gb_est520 520:gb_est521 521:gb_est522 522:gb_est523
523:gb_est524 524:gb_est525 525:gb_est526 526:gb_est527
527:gb_est528 528:gb_est529 529:gb_est530 530:gb_est531
531:gb_est532 532:gb_est533 533:gb_est534 534:gb_est535
535:gb_est536 536:gb_est537 537:gb_est538 538:gb_est539
539:gb_est540 540:gb_est541 541:gb_est542 542:gb_est543
543:gb_est544 544:gb_est545 545:gb_est546 546:gb_est547
547:gb_est548 548:gb_est549 549:gb_est550 550:gb_est551
551:gb_est552 552:gb_est553 553:gb_est554 554:gb_est555
555:gb_est556 556:gb_est557 557:gb_est558 558:gb_est559
559:gb_est560 560:gb_est561 561:gb_est562 562:gb_est563
563:gb_est564 564:gb_est565 565:gb_est566 566:gb_est567
567:gb_est568 568:gb_est569 569:gb_est570 570:gb_est571
571:gb_est572 572:gb_est573 573:gb_est574 574:gb_est575
575:gb_est576 576:gb_est577 577:gb_est578 578:gb_est579
579:gb_est580 580:gb_est581 581:gb_est582 582:gb_est583
583:gb_est584 584:gb_est585 585:gb_est586 586:gb_est587
587:gb_est588 588:gb_est589 589:gb_est590 590:gb_est591
591:gb_est592 592:gb_est593 593:gb_est594 594:gb_est595
595:gb_est596 596:gb_est597 597:gb_est598 598:gb_est599
599:gb_est600 600:gb_est601 601:gb_est602 602:gb_est603
603:gb_est604 604:gb_est605 605:gb_est606 606:gb_est607
607:gb_est608 608:gb_est609 609:gb_est610 610:gb_est611
611:gb_est612 612:gb_est613 613:gb_est614 614:gb_est615
615:gb_est616 616:gb_est617 617:gb_est618 618:gb_est619
619:gb_est620 620:gb_est621 621:gb_est622 622:gb_est623
623:gb_est624 624:gb_est625 625:gb_est626 626:gb_est627
627:gb_est628 628:gb_est629 629:gb_est630 630:gb_est631
631:gb_est632 632:gb_est633 633:gb_est634 634:gb_est635
635:gb_est636 636:gb_est637 637:gb_est638 638:gb_est639
639:gb_est640 640:gb_est641 641:gb_est642 642:gb_est643
643:gb_est644 644:gb_est645 645:gb_est646 646:gb_est647
647:gb_est648 648:gb_est649 649:gb_est650 650:gb_est651
651:gb_est652 652:gb_est653 653:gb_est654 654:gb_est655
655:gb_est656 656:gb_est657 657:gb_est658 658:gb_est659
659:gb_est660 660:gb_est661 661:gb_est662 662:gb_est663
663:gb_est664 664:gb_est665 665:gb_est666 666:gb_est667
667:gb_est668 668:gb_est669 669:gb_est670 670:gb_est671
671:gb_est672 672:gb_est673 673:gb_est674 674:gb_est675
675:gb_est676 676:gb_est677 677:gb_est678 678:gb_est679
679:gb_est680 680:gb_est681 681:gb_est682 682:gb_est683
683:gb_est684 684:gb_est685 685:gb_est686 686:gb_est687
687:gb_est688 688:gb_est689 689:gb_est690 690:gb_est691
691:gb_est692 692:gb_est693 693:gb_est694 694:gb_est695
695:gb_est696 696:gb_est697 697:gb_est698 698:gb_est699
699:gb_est700 700:gb_est701 701:gb_est702 702:gb_est703
703:gb_est704 704:gb_est705 705:gb_est706 706:gb_est707
707:gb_est708 708:gb_est709 709:gb_est710 710:gb_est711
711:gb_est712 712:gb_est713 713:gb_est714 714:gb_est715
715:gb_est716 716:gb_est717 717:gb_est718 718:gb_est719
719:gb_est720 720:gb_est721 721:gb_est722 722:gb_est723
723:gb_est724 724:gb_est725 725:gb_est726 726:gb_est727
727:gb_est728 728:gb_est729 729:gb_est730 730:gb_est731
731:gb_est732 732:gb_est733 733:gb_est734 734:gb_est735
735:gb_est736 736:gb_est737 737:gb_est738 738:gb_est739
739:gb_est740 740:gb_est741 741:gb_est742 742:gb_est743
743:gb_est744 744:gb_est745 745:gb_est746 746:gb_est747
747:gb_est748 748:gb_est749 749:gb_est750 750:gb_est751
751:gb_est752 752:gb_est753 753:gb_est754 754:gb_est755
755:gb_est756 756:gb_est757 757:gb_est758 758:gb_est759
759:gb_est760 760:gb_est761 761:gb_est762 762:gb_est763
763:gb_est764 764:gb_est765 765:gb_est766 766:gb_est767
767:gb_est768 768:gb_est769 769:gb_est770 770:gb_est771
771:gb_est772 772:gb_est773 773:gb_est774 774:gb_est775
775:gb_est776 776:gb_est777 777:gb_est778 778:gb_est779
779:gb_est780 780:gb_est781 781:gb_est782 782:gb_est783
783:gb_est784 784:gb_est785 785:gb_est786 786:gb_est787
787:gb_est788 788:gb_est789 789:gb_est790 790:gb_est791
791:gb_est792 792:gb_est793 793:gb_est794 794:gb_est795
795:gb_est796 796:gb_est797 797:gb_est798 798:gb_est799
799:gb_est800 800:gb_est801 801:gb_est802 802:gb_est803
803:gb_est804 804:gb_est805 805:gb_est806 806:gb_est807
807:gb_est808 808:gb_est809 809:gb_est810 810:gb_est811
811:gb_est812 812:gb_est813 813:gb_est814 814:gb_est815
815:gb_est816 816:gb_est817 817:gb_est818 818:gb_est819
819:gb_est820 820:gb_est821 821:gb_est822 822:gb_est823
823:gb_est824 824:gb_est825 825:gb_est826 826:gb_est827
827:gb_est828 828:gb_est829 829:gb_est830 830:gb_est831
831:gb_est832 832:gb_est833 833:gb_est834 834:gb_est835
835:gb_est836 836:gb_est837 837:gb_est838 838:gb_est839
839:gb_est840 840:gb_est841 841:gb_est842 842:gb_est843
843:gb_est844 844:gb_est845 845:gb_est846 846:gb_est847
847:gb_est848 848:gb_est849 849:gb_est850 850:gb_est851
851:gb_est852 852:gb_est853 853:gb_est854 854:gb_est855
855:gb_est856 856:gb_est857 857:gb_est858 858:gb_est859
859:gb_est860 860:gb_est861 861:gb_est862 862:gb_est863
863:gb_est864 864:gb_est865 865:gb_est866 866:gb_est867
867:gb_est868 868:gb_est869 869:gb_est870 870:gb_est871
871:gb_est872 872:gb_est873 873:gb_est874 874:gb_est875
875:gb_est876 876:gb_est877 877:gb_est878 878:gb_est879
879:gb_est880 880:gb_est881 881:gb_est882 882:gb_est883
883:gb_est884 884:gb_est885 885:gb_est886 886:gb_est887
887:gb_est888 888:gb_est889 889:gb_est890 890:gb_est891
891:gb_est892 892:gb_est893 893:gb_est894 894:gb_est895
895:gb_est896 896:gb_est897 897:gb_est898 898:gb_est899
899:gb_est900 900:gb_est901 901:gb_est902 902:gb_est903
903:gb_est904 904:gb_est905 905:gb_est906 906:gb_est907
907:gb_est908 908:gb_est909 909:gb_est910 910:gb_est911
911:gb_est912 912:gb_est913 913:gb_est914 914:gb_est915
915:gb_est916 916:gb_est917 917:gb_est918 918:gb_est919
919:gb_est920 920:gb_est921 921:gb_est922 922:gb_est923
923:gb_est924 924:gb_est925 925:gb_est926 926:gb_est927
927:gb_est928 928:gb_est929 929:gb_est930 930:gb_est931
931:gb_est932 932:gb_est933 933:gb_est934 934:gb_est935
935:gb_est936 936:gb_est937 937:gb_est938 938:gb_est939
939:gb_est940 940:gb_est941 941:gb_est942 942:gb_est943
943:gb_est944 944:gb_est945 945:gb_est946 946:gb_est947
947:gb_est948 948:gb_est949 949:gb_est950 950:gb_est951
951:gb_est952 952:gb_est953 953:gb_est954 954:gb_est955
955:gb_est956 956:gb_est957 957:gb_est958 958:gb_est959
959:gb_est960 960:gb_est961 961:gb_est962 962:gb_est963
963:gb_est964 964:gb_est965 965:gb_est966 966:gb_est967
967:gb_est968 968:gb_est969 969:gb_est970 970:gb_est971
971:gb_est972 972:gb_est973 973:gb_est974 974:gb_est975
975:gb_est976 976:gb_est977 977:gb_est978 978:gb_est979
979:gb_est980 980:gb_est981 981:gb_est982 982:gb_est983
983:gb_est984 984:gb_est985 985:gb_est986 986:gb_est987
987:gb_est988 988:gb_est989 989:gb_est990 990:gb_est991
991:gb_est992 992:gb_est993 993:gb_est994 994:gb_est995
995:gb_est996 996:gb_est997 997:gb_est998 998:gb_est999
999:gb_est1000 1000:gb_est1001 1001:gb_est1002 1002:gb_est1003
1003:gb_est1004 1004:gb_est1005 1005:gb_est1006 1006:gb_est1007
1007:gb_est1008 1008:gb_est1009 1009:gb_est1010 1010:gb_est1011
1011:gb_est1012 1012:gb_est1013 1013:gb_est1014 1014:gb_est1015
1015:gb_est1016 1016:gb_est1017 1017:gb_est1018 1018:gb_est1019
1019:gb_est1020 1020:gb_est1021 1021:gb_est1022 1022:gb_est1023
1023:gb_est1024 1024:gb_est1025 1025:gb_est1026 1026:gb_est1027
1027:gb_est1028 1028:gb_est1029 1029:gb_est1030 1030:gb_est1031
1031:gb_est1032 1032:gb_est1033 1033:gb_est1034 1034:gb_est1035
1035:gb_est1036 1036:gb_est1037 1037:gb_est1038 1038:gb_est1039
1039:gb_est1040 1040:gb_est1041 1041:gb_est1042 1042:gb_est1043
1043:gb_est1044 1044:gb_est1045 1045:gb_est1046 1046:gb_est1047
1047:gb_est1048 1048:gb_est1049 1049:gb_est1050 1050:gb_est1051
1051:gb_est1052 1052:gb_est1053 1053:gb_est1054 1054:gb_est1055
1055:gb_est1056 1056:gb_est1057 1057:gb_est1058 1058:gb_est1059
1059:gb_est1060 1060:gb_est1061 1061:gb_est1062 1062:gb_est1063
1063:gb_est1064 1064:gb_est1065 1065:gb_est1066 1066:gb_est1067
1067:gb_est1068 1068:gb_est1069 1069:gb_est1070 1070:gb_est1071
1071:gb_est1072 1072:gb_est1073 1073:gb_est1074 1074:gb_est1075
1075:gb_est1076 1076:gb_est1077 1077:gb_est1078 1078:gb_est1079
1079:gb_est1080 1080:gb_est1081 1081:gb_est1082 1082:gb_est1083
1083:gb_est1084 1084:gb_est1085 1085:gb_est1086 1086:gb_est1087
1087:gb_est1088 1088:gb_est1089 1089:gb_est1090 1090:gb_est1091
1091:gb_est1092 1092:gb_est1093 1093:gb_est1094 1094:gb_est1095
1095:gb_est1096 1096:gb_est1097 1097:gb_est1098 1098:gb_est1099
1099:gb_est1100 1100:gb_est1101 1101:gb_est1102 1102:gb_est1103
1103:gb_est1104 1104:gb_est1105 1105:gb_est1106 1106:gb_est1107
1107:gb_est1108 1108:gb_est1109 1109:gb_est1110 1110:gb_est1111
1111:gb_est1112 1112:gb_est1113 1113:gb_est1114 1114:gb_est1115
1115:gb_est1116 1116:gb_est1117 1

Best Local Similarity 73.4%; Pred. No. 4.84e-05;					
Matches 47; Conservative 0; Mismatches 17; Indels 0; Gaps 0;					
Db	31	TTTTTTT	TTTTTTGGCTTA	AAAAAACACAAAATTTT	TGACAAAGACAA 90
QY	5426	TTTTTTT	TTTGGTTTTC	CACAAAACCAAAATCAATT	TCTGCATAATCCAA 5485
Db	91	AAAG 94			
QY	5486	AAAG 5489			
RESULT	12				
LOCUS	A1128673	146 bp	mRNA	EST	05-OCT-1998
DEFINITION	qa60h01.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone	IMAGE:1691185 3', mRNA sequence.			
ACCESSION	A1128673				
NID	g3597187				
VERSION	A1128673.1	GI:3597187			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 146)				
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) On Jan 19, 1998 this sequence version replaced gi:2152106.				
Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1138 Std Error: 0.00 Seq primer: -40ml3 fwd. ET from Amersham High quality sequence stop: 93. Location/Qualifiers 1..146 /organism="Homo sapiens" /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGGCCCATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NbHL19W." /db_xref="taxon:9606" /clone=IMAGE:1691185 /clone_lib="Soares_fetal_heart_NbHH19W" /sex="unknown" /dev_stage="19 weeks" /lab_host="DH10B (ampicillin resistant)" BASE COUNT 44 a 18 c 14 g 70 t ORIGIN					
Query Match 3.0%; Score 30; DB 23; Length 146; Best Local Similarity 73.4%; Pred. No. 4.84e-05; Matches 47; Conservative 0; Mismatches 17; Indels 0; Gaps 0;					
Db	18	TTTTTTTTTTTTTTT	CTGCTTAAAAAACA	AAAAAAAAATTTTTTT	TGACAAAGACAA 77
QY	5426	TTTTTTTTTTT	GGTTTTTTTG	TCCAAAACCAAAATCAATT	TCTGCATAATCCAA 5485
Db	78	AAAG 81			

Search completed: Sat Aug 7 21:29:52 1999
Job time : 1738 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 10 11:44:07 1999; MasPar time 23.35 Seconds
Tabular output not generated. 458.057 Million cell updates/sec

Title: >US-08-287-669-19
Description: (1-503) from US08287669.pep
Perfect Score: 503
Sequence: 1 MMRQDRRSLLERNIMFSSH.....MTSRLKKFYFWPEARNASV 503

Scoring table: TABLE unitprotatable
Gap 60

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 2.996; Variance 0.689; scale 4.345

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	503	100.0	503	9	R47466	Ced-3.
2	503	100.0	503	19	R98754	Nematode Ced-3 cell d
3	485	96.4	503	9	R53288	Ced-3 (S486F).
4	482	95.8	503	9	R53287	Ced-3 (E483K).
5	476	94.6	503	9	R53279	Ced-3 (L27F).
6	466	92.6	503	9	R45327	Ced-3 mutant Y493.
7	466	92.6	503	9	R45262	Ced-3.
8	466	92.6	503	9	R45328	Ced-3 mutant P496.
9	466	92.6	503	9	R45326	Ced-3 mutant L488.
10	465	92.4	503	9	R45282	Ced-3 mutant N250.
11	465	92.4	503	9	R45319	Ced-3 mutant D414.
12	465	92.4	503	9	R45281	Ced-3 mutant I248.
13	465	92.4	503	9	R45294	Ced-3 mutant M291.
14	465	92.4	503	9	R45303	Ced-3 mutant G316.
15	465	92.4	503	9	R45279	Ced-3 mutant L246.
16	465	92.4	503	9	R45297	Ced-3 mutant D306.

17	465	92.4	503	9	R45310	Ced-3 mutant K349.	0.00e+00
18	465	92.4	503	9	R45323	Ced-3 mutant I438.	0.00e+00
19	465	92.4	503	9	R45285	Ced-3 mutant G261.	0.00e+00
20	465	92.4	503	9	R45293	Ced-3 mutant T287.	0.00e+00
21	465	92.4	503	9	R45308	Ced-3 mutant P344.	0.00e+00
22	465	92.4	503	9	R45277	Ced-3 mutant M234.	0.00e+00
23	465	92.4	503	9	R45280	Ced-3 mutant I247.	0.00e+00
24	465	92.4	503	9	R45315	Ced-3 mutant C358.	0.00e+00
25	465	92.4	503	9	R45287	Ced-3 mutant G277.	0.00e+00
26	465	92.4	503	9	R45290	Ced-3 mutant K283.	0.00e+00
27	465	92.4	503	9	R45289	Ced-3 mutant V280.	0.00e+00
28	465	92.4	503	9	R45312	Ced-3 mutant K351.	0.00e+00
29	465	92.4	503	9	R45313	Ced-3 mutant Q356.	0.00e+00
30	465	92.4	503	9	R45286	Ced-3 mutant D265.	0.00e+00
31	465	92.4	503	9	R45322	Ced-3 mutant S435.	0.00e+00
32	465	92.4	503	9	R45295	Ced-3 mutant F298.	0.00e+00
33	465	92.4	503	9	R45291	Ced-3 mutant N285.	0.00e+00
34	465	92.4	503	9	R45284	Ced-3 mutant R259.	0.00e+00
35	465	92.4	503	9	R45298	Ced-3 mutant S307.	0.00e+00
36	465	92.4	503	9	R45311	Ced-3 mutant P350.	0.00e+00
37	465	92.4	503	9	R45276	Ced-3 mutant S183.	0.00e+00
38	465	92.4	503	9	R45296	Ced-3 mutant H304.	0.00e+00
39	465	92.4	503	9	R45283	Ced-3 mutant F253.	0.00e+00
40	465	92.4	503	9	R45324	Ced-3 mutant A449.	0.00e+00
41	465	92.4	503	9	R45292	Ced-3 mutant L286.	0.00e+00
42	465	92.4	503	9	R45317	Ced-3 mutant G360.	0.00e+00
43	465	92.4	503	9	R45304	Ced-3 mutant I321.	0.00e+00
44	465	92.4	503	9	R45302	Ced-3 mutant H315.	0.00e+00
45	465	92.4	503	9	R45278	Ced-3 mutant R242.	0.00e+00
46	465	92.4	503	9	R45318	Ced-3 mutant D371.	0.00e+00
47	465	92.4	503	9	R45299	Ced-3 mutant L310.	0.00e+00
48	465	92.4	503	9	R45301	Ced-3 mutant S314.	0.00e+00
49	465	92.4	503	9	R45300	Ced-3 mutant V311.	0.00e+00
50	465	92.4	503	9	R53286	Ced-3 (A466V).	0.00e+00
51	465	92.4	503	9	R45309	Ced-3 mutant L346.	0.00e+00
52	465	92.4	503	9	R45320	Ced-3 mutant R429.	0.00e+00
53	465	92.4	503	9	R45305	Ced-3 mutant G323.	0.00e+00
54	465	92.4	503	9	R45325	Ced-3 mutant V452.	0.00e+00
55	465	92.4	503	9	R45316	Ced-3 mutant R359.	0.00e+00
56	465	92.4	503	9	R45306	Ced-3 mutant I334.	0.00e+00
57	465	92.4	503	9	R45321	Ced-3 mutant G434.	0.00e+00
58	465	92.4	503	9	R45314	Ced-3 mutant A357.	0.00e+00
59	465	92.4	503	9	R45307	Ced-3 mutant N339.	0.00e+00
60	465	92.4	503	9	R45288	Ced-3 mutant Y278.	0.00e+00
61	452	89.9	503	19	R98758	Nematode Ced-3 cell d	0.00e+00
62	452	89.9	503	9	R53285	Ced-3 (A449V).	0.00e+00
63	452	89.9	503	9	R53281	Ced-3 (G360S).	0.00e+00
64	452	89.9	503	13	R66770	Interleukin-1 beta co	0.00e+00
65	452	89.9	503	9	R53280	Ced-3 (G65R).	0.00e+00
66	427	84.9	427	9	R53284	Ced-3 (W428X).	0.00e+00
67	411	81.7	411	9	R53283	Ced-3 (Q412X).	0.00e+00
68	402	79.9	402	9	R53282	Ced-3 (Q403X).	0.00e+00
69	36	7.2	505	9	R47463	Ced-3.	1.01e-45
70	28	5.6	480	9	R47464	Ced-3.	6.29e-32
71	9	1.8	303	22	W15262	Apototic protease Mch	6.87e-02
72	9	1.8	303	22	W15247	Cysteine protease CMH	6.87e-02
73	9	1.8	341	18	R95830	Human interleukin-1-c	6.87e-02
74	8	1.6	20	27	W30019	Cytokine hCNTF AB loo	1.35e+00
75	8	1.6	20	13	R70774	CNTF neurotrophic pep	1.35e+00
76	8	1.6	20	35	W66137	Prosaposin receptor a	1.35e+00
77	8	1.6	199	38	W83335	Rabbit ciliary neurot	1.35e+00
78	8	1.6	199	2	R06556	Rabbit sciatic nerve	1.35e+00
79	8	1.6	199	1	R06233	Rabbit sciatic nerve	1.35e+00
80	8	1.6	199	7	R34429	Sequence of rabbit ci	1.35e+00
81	8	1.6	200	38	W83343	Modified ciliary neur	1.35e+00
82	8	1.6	200	38	W83334	Rat ciliary neurotrop	1.35e+00
83	8	1.6	200	38	W83344	Modified ciliary neur	1.35e+00
84	8	1.6	200	38	W83338	Modified ciliary neur	1.35e+00
85	8	1.6	200	38	W83346	Modified ciliary neur	1.35e+00
86	8	1.6	200	38	W83341	Modified ciliary neur	1.35e+00
87	8	1.6	200	32	W37931	Amino acid sequence o	1.35e+00
88	8	1.6	200	10	R53431	Human/rat chimeric ci	1.35e+00
89	8	1.6	200	10	R53426	Human/rat chimeric ci	1.35e+00

90	1.6	200 10	R53428	Human/rat chimeric ci	1.35e+00	163	1.2	198 11	R59841	ApoE4L protease.	3.00e+02
91	1.6	200 10	R53423	Human/rat chimeric ci	1.35e+00	164	1.2	199 31	W48651	Puromycin acetyl tran	3.00e+02
92	1.6	200 2	R11617	Rat ciliary neurotrop	1.35e+00	165	1.2	200 38	W83345	Modified ciliary neur	3.00e+02
93	1.6	200 10	R53429	Human/rat chimeric ci	1.35e+00	166	1.2	200 10	R53430	Human/rat chimeric ci	3.00e+02
94	1.6	200 7	R34430	Sequence of rat cilia	1.35e+00	167	1.2	203 28	W39208	Human ICE LAP-6 amino	3.00e+02
95	1.6	821 7	R35451	Mouse eps8.	1.35e+00	168	1.2	205 28	W38184	Arabidopsis SCARECROW	3.00e+02
96	1.4	15 16	R88457	erbB-3 peptide (aa 11	2.25e+01	169	1.2	216 15	R75909	Human olfactory recep	3.00e+02
97	1.4	15 36	W69411	ErbB-3 glycoprotein i	2.25e+01	170	1.2	216 15	R75910	Human olfactory recep	3.00e+02
98	1.4	101 14	R77251	Guinea pig G protein	2.25e+01	171	1.2	227 31	W53835	Pseudomonas xcpW secr	3.00e+02
99	1.4	134 14	R72393	Agmatellum quadruplic	2.25e+01	172	1.2	230 23	W08430	Rapamycin-dependent a	3.00e+02
100	1.4	135 29	W55413	H. pylori ORF 29ep201	2.25e+01	173	1.2	232 37	W75092	Human secreted protei	3.00e+02
101	1.4	162 13	R67983	23kDa antigen encoded	2.25e+01	174	1.2	235 31	W27951	Amino acid sequence o	3.00e+02
102	1.4	164 38	W82399	Human G3BP protein.	2.25e+01	175	1.2	236 10	R54225	L.lactis alpha-acetol	3.00e+02
103	1.4	207 7	R36389	Urease subunit from B	2.25e+01	176	1.2	248 37	W73162	Mouse amphiregulin.	3.00e+02
104	1.4	252 23	W11617	Epstein-Barr virus p2	2.25e+01	177	1.2	249 29	W38358	Apoptosis associated	3.00e+02
105	1.4	253 22	W15263	Cysteine protease Mch	2.25e+01	178	1.2	254 12	R64190	Human 4-1BB-L polypep	3.00e+02
106	1.4	364 19	R88751	Human Ty protease, re	2.25e+01	179	1.2	254 25	W26657	Human 4-1BB ligand.	3.00e+02
107	1.4	372 20	W08432	Human interleukin-1 c	2.25e+01	180	1.2	261 24	W11898	MACH isoform beta4.	3.00e+02
108	1.4	376 19	R99254	Cytoplasmic antiprote	2.25e+01	181	1.2	270 8	R42817	Class II AP endonucle	3.00e+02
109	1.4	418 15	R83098	Interleukin-1-beta co	2.25e+01	182	1.2	275 15	R84563	Human prostacyclin-sy	3.00e+02
110	1.4	567 38	W88788	Polypeptide fragment	2.25e+01	183	1.2	277 19	W00677	Pro-Yama.	3.00e+02
111	1.4	667 29	W55656	H. pylori ORF 09cp610	2.25e+01	184	1.2	277 30	W41688	Amino acid sequence o	3.00e+02
112	1.4	668 30	W55738	H. pylori ORF 01ce610	2.25e+01	185	1.2	277 34	W48945	Mutant human apopain	3.00e+02
113	1.4	672 37	W73026	Helicobacter pylori 7	2.25e+01	186	1.2	277 34	W48937	Mutant human apopain	3.00e+02
114	1.4	1342 3	R13833	HER-3 epithelial grow	2.25e+01	187	1.2	277 30	W47089	Rat interleukin-1 bet	3.00e+02
115	1.4	1342 16	R88453	erbB-3 polypeptide.	2.25e+01	188	1.2	277 18	R95831	Human interleukin-I-c	3.00e+02
116	1.4	1342 36	W69406	ERBB-3 glycoprotein c	2.25e+01	189	1.2	277 24	W11893	MACH isoform beta3.	3.00e+02
117	1.4	1343 3	R12608	EGFR-R erbB-3 clone E	2.25e+01	190	1.2	277 21	W00372	Apopain CPP32beta pro	3.00e+02
118	1.4	2161 33	W63137	Human calcium channel	2.25e+01	191	1.2	277 21	W16600	Apopain CPP32a proenz	3.00e+02
119	1.4	2161 14	R71002	Human neuronal calciu	2.25e+01	192	1.2	278 37	W74863	Human secreted protei	3.00e+02
120	1.4	2161 14	R71001	Human neuronal calciu	2.25e+01	193	1.2	283 33	W49769	Amino acid sequence o	3.00e+02
121	1.4	2161 6	R33545	Sequence of the alpha	2.25e+01	194	1.2	283 21	W05302	D-aminotransferase.	3.00e+02
122	1.4	2161 33	W63149	Human calcium channel	2.25e+01	195	1.2	284 34	W75860	Human secretory prote	3.00e+02
123	1.4	3011 16	R95021	Hepatitis GB virus (H	2.25e+01	196	1.2	290 10	R51273	SSO1 gene product.	3.00e+02
124	1.2	8 34	W78520	SH2 domain binding in	3.00e+02	197	1.2	292 19	W00641	Infectious laryngotra	3.00e+02
125	1.2	16 36	W71727	Nck middle SH3 domain	3.00e+02	198	1.2	292 21	W06793	ILTV unique short reg	3.00e+02
126	1.2	20 27	W25350	Synthetic lytic pepti	3.00e+02	199	1.2	294 37	W80486	Islet cell antibody a	3.00e+02
127	1.2	20 21	W16975	Lytic peptide used in	3.00e+02	200	1.2	297 19	W02708	G-protein coupled hum	3.00e+02
128	1.2	23 18	R89981	Synthetic lytic pepti	3.00e+02	201	1.2	297 16	R48736	G-protein coupled hum	3.00e+02
129	1.2	23 18	R92424	Lytic peptide used in	3.00e+02	202	1.2	297 16	R48737	G-protein coupled hum	3.00e+02
130	1.2	27 18	R89982	Synthetic lytic pepti	3.00e+02	203	1.2	297 19	W02709	G-protein coupled hum	3.00e+02
131	1.2	27 18	R92425	Lytic peptide used in	3.00e+02	204	1.2	298 2	P70845	Sequence of human res	3.00e+02
132	1.2	28 39	W67932	Fragment of human sec	3.00e+02	205	1.2	298 29	W47605	HRSV glycoprotein G.	3.00e+02
133	1.2	38 7	R36838	Domain 2 from Tn554 T	3.00e+02	206	1.2	298 5	R39286	Respiratory syncytial	3.00e+02
134	1.2	46 4	R22023	Prothrombin leader fo	3.00e+02	207	1.2	298 8	R25302	HRSV glycoprotein G (3.00e+02
135	1.2	47 4	R22024	Prothrombin leader fo	3.00e+02	208	1.2	298 7	R35438	B. burgdorferi strain	3.00e+02
136	1.2	52 34	W77642	Staphylococcus aureus	3.00e+02	209	1.2	311 16	R92115	Human ApoE4Lx2.	3.00e+02
137	1.2	60 31	W57073	Measles virus haemagg	3.00e+02	210	1.2	311 11	R59843	ApoE4Lx2 protease.	3.00e+02
138	1.2	60 31	W57087	Measles virus haemagg	3.00e+02	211	1.2	321 19	W02674	G-protein coupled Dro	3.00e+02
139	1.2	64 31	W28057	Staphylococcus aureus	3.00e+02	212	1.2	321 16	R48702	G-protein coupled Dro	3.00e+02
140	1.2	73 8	R41477	Peptide derived from	3.00e+02	213	1.2	324 29	W55546	H. pylori ORF 12ap103	3.00e+02
141	1.2	78 14	R74267	SSAL.	3.00e+02	214	1.2	325 33	W59645	Amino acid sequence o	3.00e+02
142	1.2	78 14	R74245	SSAL2 Inf A-HA-H1N1.	3.00e+02	215	1.2	334 10	R53644	c-jun gene gene produ	3.00e+02
143	1.2	79 35	W72028	HSV-2 strain SB5 Cont	3.00e+02	216	1.2	334 36	W72171	HSV-2 strain SB5 Cont	3.00e+02
144	1.2	80 32	W49763	Human D2-2 protein as	3.00e+02	217	1.2	334 36	W72090	HSV-2 strain SB5 Cont	3.00e+02
145	1.2	87 17	R90990	Human Mat-8 polypepti	3.00e+02	218	1.2	334 35	W72005	HSV-2 strain SB5 Cont	3.00e+02
146	1.2	95 29	W55281	H. pylori ORF 12ap103	3.00e+02	219	1.2	341 16	R48727	G-protein coupled hum	3.00e+02
147	1.2	95 36	W74972	Human secreted protei	3.00e+02	220	1.2	341 19	W02699	G-protein coupled hum	3.00e+02
148	1.2	98 9	R48598	Sequence of murine IP	3.00e+02	221	1.2	344 10	R54222	L.lactis branched ami	3.00e+02
149	1.2	100 25	W17674	FRB domain of FRAP pr	3.00e+02	222	1.2	350 32	W60244	Amino acid of the spe	3.00e+02
150	1.2	110 36	W74993	Human secreted protei	3.00e+02	223	1.2	352 34	W44931	Soybean seed coat per	3.00e+02
151	1.2	111 36	W80630	S. pneumoniae spermid	3.00e+02	224	1.2	352 37	W73138	Soybean seed coat per	3.00e+02
152	1.2	111 7	R34282	Human gamma-signalin	3.00e+02	225	1.2	354 24	W25926	Xenopus melatonin rec	3.00e+02
153	1.2	121 24	W18106	G-protein coupled rec	3.00e+02	226	1.2	354 24	W25927	Xenopus melatonin rec	3.00e+02
154	1.2	124 19	W02984	G-protein coupled rec	3.00e+02	227	1.2	355 21	W14796	Mangosteen Class I ac	3.00e+02
155	1.2	124 16	R50792	G-protein coupled rec	3.00e+02	228	1.2	355 20	W09299	Mangosteen class I ac	3.00e+02
156	1.2	130 23	W08428	PCANTAB-AP-FRAP(2015-	3.00e+02	229	1.2	356 7	R37715	Carminomycin 4-O-meth	3.00e+02
157	1.2	153 31	W54384	Actinomadura hibisca	3.00e+02	230	1.2	362 13	R74150	Brassica acyl thioest	3.00e+02
158	1.2	155 2	R12107	D.discoideum nucleosi	3.00e+02	231	1.2	362 29	W44333	Brassica campestris l	3.00e+02
159	1.2	162 28	W29461	Mad-related human JV1	3.00e+02	232	1.2	362 10	R54948	Brassica campestris l	3.00e+02
160	1.2	162 19	W04234	Mouse RAP1 partial s	3.00e+02	233	1.2	362 6	R29172	Brassica thioesterase	3.00e+02
161	1.2	195 30	W42428	Glucuronid repressor	3.00e+02	234	1.2	364 33	W57830	Rat pancreatic beta-c	3.00e+02
162	1.2	198 16	R92113	Human ApoE4L.	3.00e+02	235	1.2	370 19	W02669	G-protein coupled rat	3.00e+02

236	6	1.2	370	16	R48697	G-protein coupled rat	3.00e+02	309	6	1.2	483	23	W21639	Grapevine leafroll vi	3.00e+02
237	6	1.2	372	4	R24027	Sequence of mouse lym	3.00e+02	310	6	1.2	490	23	W10209	Full length endoglyco	3.00e+02
238	6	1.2	372	14	R76507	Mouse LHR.	3.00e+02	311	6	1.2	491	10	R51107	Wheat acetyl-CoA carb	3.00e+02
239	6	1.2	372	4	R22803	Murine lymphocyte hom	3.00e+02	312	6	1.2	496	23	W19584	Mouse apoptosis inhib	3.00e+02
240	6	1.2	372	7	R37961	Mouse lymphocyte Homi	3.00e+02	313	6	1.2	496	15	R70647	Azoospermia factor YR	3.00e+02
241	6	1.2	372	8	R38909	MLHR.	3.00e+02	314	6	1.2	496	35	W69297	Murine XIAP protein.	3.00e+02
242	6	1.2	372	18	R98107	Mouse lymphocyte cell	3.00e+02	315	6	1.2	496	27	W27391	Mch5 protein.	3.00e+02
243	6	1.2	372	3	R12470	Sequence encoding mur	3.00e+02	316	6	1.2	496	23	W19745	Mouse inhibitor of ap	3.00e+02
244	6	1.2	372	37	W73265	Mouse lymphocyte homi	3.00e+02	317	6	1.2	499	18	W03713	Human alpha-1C2 adren	3.00e+02
245	6	1.2	372	14	R83051	Mouse LHR.	3.00e+02	318	6	1.2	500	15	R84561	Human prostacyclin-sy	3.00e+02
246	6	1.2	372	1	R06242	Homing receptor unit	3.00e+02	319	6	1.2	502	29	W55550	H. pylori ORF 01ce211	3.00e+02
247	6	1.2	373	18	R98461	Murine ICE-ced-3 homo	3.00e+02	320	6	1.2	502	29	W55452	H. pylori ORF 02ae116	3.00e+02
248	6	1.2	373	31	W56031	Mouse ICH-3.	3.00e+02	321	6	1.2	515	18	R90041	Human/rat hybrid alph	3.00e+02
249	6	1.2	373	13	R66767	Murine interleukin-1	3.00e+02	322	6	1.2	515	37	W77106	Rat alpha18-adrenergi	3.00e+02
250	6	1.2	374	31	W37782	Mus musculus lymphocy	3.00e+02	323	6	1.2	515	12	R70995	Human/rat alpha-1B ad	3.00e+02
251	6	1.2	375	16	R48696	G-protein coupled cow	3.00e+02	324	6	1.2	520	14	R85943	Alpha-1B adrenergic r	3.00e+02
252	6	1.2	375	19	W02668	G-protein coupled cow	3.00e+02	325	6	1.2	520	10	R52831	Sequence of human alp	3.00e+02
253	6	1.2	376	10	R54947	Cuphea hookeriana thi	3.00e+02	326	6	1.2	520	33	W55123	Streptococcus pneumon	3.00e+02
254	6	1.2	377	19	W00213	Human Ich-2 protein.	3.00e+02	327	6	1.2	520	10	R53072	Alpha 1b adrenergic r	3.00e+02
255	6	1.2	377	15	R84740	Interleukin-1-beta co	3.00e+02	328	6	1.2	525	13	R78521	NPM/ALK fusion protei	3.00e+02
256	6	1.2	377	19	R88750	Human Tx protease, re	3.00e+02	329	6	1.2	535	33	W57570	Homo sapiens B258 seq	3.00e+02
257	6	1.2	377	15	R90702	Interleukin-1-beta co	3.00e+02	330	6	1.2	544	35	W71203	Protein encoded by OR	3.00e+02
258	6	1.2	377	36	W76330	Interleukin-1 beta co	3.00e+02	331	6	1.2	547	23	W22707	N-terminal truncated	3.00e+02
259	6	1.2	382	7	R39224	Nucleocapsid protein	3.00e+02	332	6	1.2	548	36	W72083	HSV-2 strain SB5 Cont	3.00e+02
260	6	1.2	384	9	R47236	Wild-type Feline Herp	3.00e+02	333	6	1.2	548	36	W72156	HSV-2 strain SB5 Cont	3.00e+02
261	6	1.2	397	1	P40063	Sequence of aminotran	3.00e+02	334	6	1.2	548	37	W80487	Islet cell antibody a	3.00e+02
262	6	1.2	397	1	P80666	Protein encoded by ty	3.00e+02	335	6	1.2	548	36	W72252	HSV-2 strain SB5 Cont	3.00e+02
263	6	1.2	398	33	W50281	Canine herpes virus O	3.00e+02	336	6	1.2	553	23	W22706	Arginyl-tRNA syntheta	3.00e+02
264	6	1.2	399	5	R27511	Human R3BP.	3.00e+02	337	6	1.2	556	36	W71738	Human 3-phosphoinosit	3.00e+02
265	6	1.2	407	16	R79445	Human pancreatic beta	3.00e+02	338	6	1.2	562	2	R23651	Tak long fibre protei	3.00e+02
266	6	1.2	407	33	W57829	Human pancreatic beta	3.00e+02	339	6	1.2	562	3	R12575	Ad41 long fibre prote	3.00e+02
267	6	1.2	415	13	R69521	Rat corticotropin rel	3.00e+02	340	6	1.2	566	13	R66609	Human cdc25B gene pro	3.00e+02
268	6	1.2	415	17	R97292	Rat CRF receptor.	3.00e+02	341	6	1.2	566	31	W59136	Human cdc25B protein.	3.00e+02
269	6	1.2	415	13	R69519	Human pituitary corti	3.00e+02	342	6	1.2	566	2	R08259	Haemagglutinin.	3.00e+02
270	6	1.2	415	18	W00159	Human corticotropin r	3.00e+02	343	6	1.2	566	34	W68405	SIV strain H1N1 haema	3.00e+02
271	6	1.2	415	17	R97290	Human CRF receptor CR	3.00e+02	344	6	1.2	566	28	W35317	Human cdc25B cell cyc	3.00e+02
272	6	1.2	415	17	R97294	Mouse CRF RAL recepto	3.00e+02	345	6	1.2	566	32	W37987	Amino acid sequence f	3.00e+02
273	6	1.2	416	28	W39209	Human ICE LAP-6 polyp	3.00e+02	346	6	1.2	566	18	R98213	Human CDC25B phosphat	3.00e+02
274	6	1.2	417	30	W55706	H. pylori ORF 12apl03	3.00e+02	347	6	1.2	566	7	R37494	cdc25B.	3.00e+02
275	6	1.2	419	24	W26465	Mycobacterial seryl-t	3.00e+02	348	6	1.2	568	1	R04240	Gamma-glutamyl transp	3.00e+02
276	6	1.2	419	15	R70648	Azoospermia factor YR	3.00e+02	349	6	1.2	568	30	W53451	Mouse gamma-glutamylt	3.00e+02
277	6	1.2	420	17	R88409	High-affinity melaton	3.00e+02	350	6	1.2	570	14	R78142	OP-sensitiv esterase	3.00e+02
278	6	1.2	424	10	R53498	Marmoset ZP3.	3.00e+02	351	6	1.2	570	22	W17766	Malathion carboxylest	3.00e+02
279	6	1.2	425	33	W57828	Rat pancreatic beta-c	3.00e+02	352	6	1.2	570	22	W17765	Malathion carboxylest	3.00e+02
280	6	1.2	425	16	R79447	Rat pancreatic beta c	3.00e+02	353	6	1.2	570	22	W17768	Malathion carboxylest	3.00e+02
281	6	1.2	428	23	W20053	S. aureus seryl-trNA	3.00e+02	354	6	1.2	571	30	W55997	Protein SEQ ID NO:227	3.00e+02
282	6	1.2	429	18	W03714	Human alpha-1C3 adren	3.00e+02	355	6	1.2	572	22	W01670	Influenza A/Texas/36/	3.00e+02
283	6	1.2	436	33	W69513	Mouse Pax6 protein SE	3.00e+02	356	6	1.2	572	13	R78520	Partial ALK protein.	3.00e+02
284	6	1.2	436	33	W69509	Mouse Pax6 protein.	3.00e+02	357	6	1.2	580	29	W47183	Pyruvate decarboxylas	3.00e+02
285	6	1.2	451	12	R63088	Breast tumor kinase,	3.00e+02	358	6	1.2	588	9	R47813	Sequence encoded by t	3.00e+02
286	6	1.2	459	17	R77432	Moraxella catarrhalis	3.00e+02	359	6	1.2	601	4	R21931	D.melanogaster octopa	3.00e+02
287	6	1.2	461	23	W10210	Mature endoglycoceram	3.00e+02	360	6	1.2	603	13	R74620	Human lung tumour Pol	3.00e+02
288	6	1.2	464	39	W86008	Helicobacter pylori a	3.00e+02	361	6	1.2	607	22	W09422	Banana polyphenol oxi	3.00e+02
289	6	1.2	464	24	W11895	MACH isoform alpha2.	3.00e+02	362	6	1.2	614	15	R82630	70K autoantigen, part	3.00e+02
290	6	1.2	466	18	R90037	Human alpha-1C adrene	3.00e+02	363	6	1.2	642	11	R60619	Carcinoembryonic anti	3.00e+02
291	6	1.2	466	12	R58680	Human alpha-1C adrene	3.00e+02	364	6	1.2	644	9	R44529	pNPX30 xylanase.	3.00e+02
292	6	1.2	466	14	R85944	Alpha-1C adrenergic r	3.00e+02	365	6	1.2	649	38	W90005	Expressed antigen for	3.00e+02
293	6	1.2	466	12	R58679	Human alpha-1C adrene	3.00e+02	366	6	1.2	663	18	R98519	Immunogenic carcinoem	3.00e+02
294	6	1.2	466	10	R53073	Alpha 1c adrenergic r	3.00e+02	367	6	1.2	668	38	W87488	S. cerevisiae TIH2 po	3.00e+02
295	6	1.2	466	14	R73603	Human alpha 1c adrene	3.00e+02	368	6	1.2	668	31	W54168	S.cerevisiae TIH2 pro	3.00e+02
296	6	1.2	466	14	R73604	Human alpha 1c adrene	3.00e+02	369	6	1.2	668	14	R81315	Yeast TIH2.	3.00e+02
297	6	1.2	466	10	R52832	Sequence of human alp	3.00e+02	370	6	1.2	691	3	R13502	HSF.	3.00e+02
298	6	1.2	466	18	R90038	Human alpha-1C adrene	3.00e+02	371	6	1.2	691	5	R24947	Sequence encoded by D	3.00e+02
299	6	1.2	468	2	R12109	sod2 gene-encoded sod	3.00e+02	372	6	1.2	697	29	W46517	Saccharomyces cerevis	3.00e+02
300	6	1.2	476	13	R74970	Murine growth differe	3.00e+02	373	6	1.2	698	23	W22844	Human carcinoembryoni	3.00e+02
301	6	1.2	476	9	R47587	Rat bone formation-in	3.00e+02	374	6	1.2	698	14	R65168	Carcinoembryonic anti	3.00e+02
302	6	1.2	477	1	P90166	Herpes Simplex virus-	3.00e+02	375	6	1.2	698	1	P81229	Carcinoembryonic anti	3.00e+02
303	6	1.2	478	9	R47586	Human bone formation-	3.00e+02	376	6	1.2	702	20	W06872	Carcinoembryonic anti	3.00e+02
304	6	1.2	478	20	W06539	Human bone morphogeni	3.00e+02	377	6	1.2	702	38	W83137	Carcinoembryonic anti	3.00e+02
305	6	1.2	479	36	W75963	Human Caspase-8.	3.00e+02	378	6	1.2	702	1	P81222	CEA protein.	3.00e+02
306	6	1.2	479	28	W31523	Interleukin-1 beta co	3.00e+02	379	6	1.2	702	1	P94014	Carcinoembryonic anti	3.00e+02
307	6	1.2	479	24	W11891	MACH isoform alphas1.	3.00e+02	380	6	1.2	702	2	P93999	Amino acid sequence o	3.00e+02
308	6	1.2	480	36	W72251	HSV-2 strain SB5 Cont	3.00e+02	381	6	1.2	702	9	R54713	Carcinoembryonic anti	3.00e+02

382	6	1.2	718 39	W84067	Drosophila mitofusin	3.00e+02	455	6	1.2	1261 25	W10040	Mad binding protein,	3.00e+02
383	6	1.2	728 15	R75789	Murine predicted MLH1	3.00e+02	456	6	1.2	1264 7	R36730	Ubiquitin-specific pr	3.00e+02
384	6	1.2	734 18	W00182	Carcinoembryonic anti	3.00e+02	457	6	1.2	1289 19	R97370	Phage T4 tail fibre p	3.00e+02
385	6	1.2	737 1	R05268	Amino acid sequence o	3.00e+02	458	6	1.2	1539 21	W13905	Human SMCY protein.	3.00e+02
386	6	1.2	739 16	R77501	Malate synthase.	3.00e+02	459	6	1.2	1671 1	R06341	Signal peptide and fi	3.00e+02
387	6	1.2	740 34	W61533	Human Fas-binding pro	3.00e+02	460	6	1.2	1686 34	W70991	Human class II p13 ki	3.00e+02
388	6	1.2	748 22	W13384	Human protein ubiquit	3.00e+02	461	6	1.2	1726 29	W38756	Phosphatidyl inositol	3.00e+02
389	6	1.2	748 29	W55511	H. pylori ORF 05ae302	3.00e+02	462	6	1.2	1786 35	W50893	Human laminin B1 chai	3.00e+02
390	6	1.2	748 30	W55725	H. pylori ORF 05ae302	3.00e+02	463	6	1.2	1891 26	W23720	Platenolide synthase	3.00e+02
391	6	1.2	751 4	R23582	Branching enzyme.	3.00e+02	464	6	1.2	1891 26	W22610	Platenolide synthase	3.00e+02
392	6	1.2	758 37	W81771	Tobacco CHLD protein.	3.00e+02	465	6	1.2	1996 26	W22607	Platenolide synthase	3.00e+02
393	6	1.2	761 39	W70899	Protein encoded by tu	3.00e+02	466	6	1.2	1996 26	W23717	Platenolide synthase	3.00e+02
394	6	1.2	766 28	W37352	Psbc (WbpC) protein i	3.00e+02	467	6	1.2	2101 9	R47173	Sequence of the inner	3.00e+02
395	6	1.2	774 3	P60616	Cephalosporin C acycl	3.00e+02	468	6	1.2	2172 35	W70409	Wheat acetyl-CoA carb	3.00e+02
396	6	1.2	781 35	W69369	Modified retinoblasto	3.00e+02	469	6	1.2	2192 23	W21732	LexA/NuMA fusion prot	3.00e+02
397	6	1.2	806 13	R65495	Marek's disease virus	3.00e+02	470	6	1.2	2257 35	W70407	Triticum aestivum ace	3.00e+02
398	6	1.2	816 35	W71355	Protein sequence of t	3.00e+02	471	6	1.2	2257 20	W05209	Wheat acetyl-CoA carb	3.00e+02
399	6	1.2	816 12	R63564	Retinoblastoma 94kd t	3.00e+02	472	6	1.2	2260 20	W05212	Wheat acetyl-CoA carb	3.00e+02
400	6	1.2	816 1	P82112	Human retinoblastoma	3.00e+02	473	6	1.2	2272 23	W21731	GAL4/HA/NuMA fusion p	3.00e+02
401	6	1.2	816 11	R58568	Human retinoblastoma	3.00e+02	474	6	1.2	2549 32	W56027	FRAP (tor1) protein.	3.00e+02
402	6	1.2	822 32	W56490	Zea mays starch branc	3.00e+02	475	6	1.2	2549 16	R81730	Sirolimus effector pr	3.00e+02
403	6	1.2	830 37	W80314	CIITA (class II trans	3.00e+02	476	6	1.2	2549 19	W04235	Human RAP1.	3.00e+02
404	6	1.2	832 35	W69368	Modified retinoblasto	3.00e+02	477	6	1.2	2721 2	P70647	Sequence of N-termina	3.00e+02
405	6	1.2	833 34	W42636	Protein sequence that	3.00e+02	478	6	1.2	2893 35	W71556	Helicobacter polypept	3.00e+02
406	6	1.2	833 34	W42637	Protein sequence that	3.00e+02	479	6	1.2	3038 14	R74171	Aspergillus terreus t	3.00e+02
407	6	1.2	844 23	W19213	Corn starch branching	3.00e+02	480	6	1.2	3066 27	W36178	Murine Ataxia-telang	3.00e+02
408	6	1.2	851 35	W69367	Modified retinoblasto	3.00e+02	481	6	1.2	3075 35	W50892	Human laminin A chain	3.00e+02
409	6	1.2	851 35	W69370	Modified retinoblasto	3.00e+02	482	6	1.2	3084 2	P94758	Sequence of mouse lam	3.00e+02
410	6	1.2	869 35	W69374	Modified retinoblasto	3.00e+02	483	6	1.2	3084 35	W50891	Mouse laminin A chain	3.00e+02
411	6	1.2	871 35	W69375	Modified retinoblasto	3.00e+02	484	6	1.2	3144 29	W44742	Human huntingtin prot	3.00e+02
412	6	1.2	874 35	W69366	Modified retinoblasto	3.00e+02	485	6	1.2	3144 11	R58777	Protein encoded by Hu	3.00e+02
413	6	1.2	874 20	W07600	Human G-protein recep	3.00e+02	486	6	1.2	3144 22	W09871	Human huntingtin.	3.00e+02
414	6	1.2	884 35	W71375	Yeast telomerase cata	3.00e+02	487	6	1.2	3144 26	W36887	Previously undescribe	3.00e+02
415	6	1.2	884 33	W59666	Amino acid sequence o	3.00e+02	488	6	1.2	3164 16	R94345	Hepatitis GB virus (H	3.00e+02
416	6	1.2	884 33	W59778	Amino acid sequence o	3.00e+02	489	6	1.2	3413 30	W52849	A. mediterranei rifam	3.00e+02
417	6	1.2	887 19	R99800	NTII-1 nerve protein,	3.00e+02	490	6	1.2	3722 2	R10145	Cephalosporin antibio	3.00e+02
418	6	1.2	895 35	W69365	Modified retinoblasto	3.00e+02	491	6	1.2	4536 28	W41262	Apolipoprotein B-100.	3.00e+02
419	6	1.2	897 35	W69371	Modified retinoblasto	3.00e+02	492	6	1.2	4545 26	W22611	Hybrid smgG/tylg ORF1	3.00e+02
420	6	1.2	916 9	R45779	PTPase PTP35 #1.	3.00e+02	493	6	1.2	4550 26	W23716	Platenolide synthase	3.00e+02
421	6	1.2	928 35	W69376	Modified retinoblasto	3.00e+02	494	6	1.2	4550 26	W22606	Platenolide synthase	3.00e+02
422	6	1.2	928 1	R05305	Cancer supressing gen	3.00e+02	495	6	1.2	4572 30	W52845	A. mediterranei rifam	3.00e+02
423	6	1.2	928 35	W69364	Modified retinoblasto	3.00e+02	496	6	1.2	4630 23	W19629	Streptomyces venezuel	3.00e+02
424	6	1.2	928 13	R71680	Retinoblastoma p110R	3.00e+02	497	6	1.2	5069 30	W52846	A. mediterranei rifam	3.00e+02
425	6	1.2	928 21	W09411	Retinoblastoma suscep	3.00e+02	498	5	1.0	8 21	W05359	Residues 363-370 of p	2.99e+03
426	6	1.2	928 14	R74271	Retinoblastoma tumour	3.00e+02	499	5	1.0	10 38	W82434	TOMY viral replicatio	2.99e+03
427	6	1.2	928 13	R71681	Recombinant p110RB p	3.00e+02	500	5	1.0	10 2	R07402	Immunogenic peptide #	2.99e+03
428	6	1.2	928 37	W62465	Retinoblastoma protei	3.00e+02	501	5	1.0	14 38	R83410	Rat Saposin C derived	2.99e+03
429	6	1.2	928 1	R06289	Predicted retinoblast	3.00e+02	502	5	1.0	20 21	W05354	Residues 363-382 of p	2.99e+03
430	6	1.2	928 35	W71354	Protein sequence of t	3.00e+02	503	5	1.0	20 13	R71650	CS Region II+ mimetic	2.99e+03
431	6	1.2	928 7	R36534	Retinoblastoma (RB) p	3.00e+02	504	5	1.0	20 13	R71651	CS Region II+ mimetic	2.99e+03
432	6	1.2	930 39	W75426	Human aggregcan degrad	3.00e+02	505	5	1.0	21 38	W88915	Polypeptide fragment	2.99e+03
433	6	1.2	961 9	R49039	PTPase PTP35 #2.	3.00e+02	506	5	1.0	26 39	W70855	Human Zcytor5 epitope	2.99e+03
434	6	1.2	970 1	P90599	Human retinoblastoma.	3.00e+02	507	5	1.0	27 7	R35430	FMD peptide A.	2.99e+03
435	6	1.2	982 3	R13320	Murine Natural Killer	3.00e+02	508	5	1.0	28 15	R89086	Human growth hormone	2.99e+03
436	6	1.2	987 31	W57057	Class II trans activa	3.00e+02	509	5	1.0	29 15	R89085	Human growth hormone	2.99e+03
437	6	1.2	1023 3	R13319	Partial Human Natural	3.00e+02	510	5	1.0	29 3	P50657	Sequence of new pepti	2.99e+03
438	6	1.2	1028 34	W29667	Homo sapiens DL185_1	3.00e+02	511	5	1.0	29 15	R89089	Human growth hormone	2.99e+03
439	6	1.2	1040 33	W55121	Streptococcus pneumon	3.00e+02	512	5	1.0	29 3	P60310	Sequence of peptide a	2.99e+03
440	6	1.2	1093 12	R66460	AF-17 protein.	3.00e+02	513	5	1.0	34 38	W81891	Synthetic pTH and pTH	2.99e+03
441	6	1.2	1100 3	R15783	B.thuringiensis toxin	3.00e+02	514	5	1.0	38 11	R65150	Specific TSAR 9-6 gen	2.99e+03
442	6	1.2	1106 37	W80312	CIITA (class II trans	3.00e+02	515	5	1.0	47 33	W62259	Subtilase TAPROK frag	2.99e+03
443	6	1.2	1129 32	W50144	Oat phytochrome A apo	3.00e+02	516	5	1.0	48 39	W70854	Human Zcytor5 epitope	2.99e+03
444	6	1.2	1130 14	R72452	Class II transactivat	3.00e+02	517	5	1.0	49 38	W88849	Polypeptide fragment	2.99e+03
445	6	1.2	1130 31	W57056	Class II trans activa	3.00e+02	518	5	1.0	50 39	W89328	Neisseria meningitidi	2.99e+03
446	6	1.2	1130 16	R81569	Class II transactivat	3.00e+02	519	5	1.0	50 3	R13350	Pl192 HCV antigen (11	2.99e+03
447	6	1.2	1130 37	W80313	CIITA (class II trans	3.00e+02	520	5	1.0	52 3	R12876	Non-collagenous bone	2.99e+03
448	6	1.2	1140 16	R81731	GST-SEP fusion protei	3.00e+02	521	5	1.0	54 39	W87855	LacZ alpha peptide (a	2.99e+03
449	6	1.2	1141 13	R66391	Human SREBP-2.	3.00e+02	522	5	1.0	54 39	W87854	LacZ alpha peptide (a	2.99e+03
450	6	1.2	1207 37	W80311	CIITA (class II trans	3.00e+02	523	5	1.0	60 39	W87788	Beta-galactosidase al	2.99e+03
451	6	1.2	1214 16	R80097	Black widow spider de	3.00e+02	524	5	1.0	61 38	W88935	Sequence ID #635 from	2.99e+03
452	6	1.2	1228 25	W22862	Bacillus stearothermo	3.00e+02	525	5	1.0	68 38	W88686	Secreted protein enco	2.99e+03
453	6	1.2	1228 17	R77673	S-layer protein encod	3.00e+02	526	5	1.0	74 2	R11780	Enzyme donor polypept	2.99e+03
454	6	1.2	1253 25	W10038	Mad binding protein,	3.00e+02	527	5	1.0	89 2	R11779	Enzyme donor polypept	2.99e+03

528	5	1.0	90	2	R11783	Enzyme donor polypept	2.99e+03	601	216	38	W89013	Polypeptide fragment	2.99e+03
529	5	1.0	92	39	W95477	Mouse derived RT3 pha	2.99e+03	602	216	28	W34203	Streptomyces non-memb	2.99e+03
530	5	1.0	100	38	W89015	Polypeptide fragment	2.99e+03	603	219	38	W88556	Secreted protein enco	2.99e+03
531	5	1.0	101	7	R34268	Mutated human VH6 gen	2.99e+03	604	219	22	W20650	H. pylori surface or	2.99e+03
532	5	1.0	104	39	W95223	PT toxin beta-subunit	2.99e+03	605	224	17	R94615	MAGE protein.	2.99e+03
533	5	1.0	105	38	W83132	Nicotiana excelsior t	2.99e+03	606	225	2	P70061	Sequence of human gro	2.99e+03
534	5	1.0	105	10	R53961	Hepatocyte growth fac	2.99e+03	607	231	13	R70000	OPDE 29 kDa subunit.	2.99e+03
535	5	1.0	107	19	W01299	VL region of HIV neut	2.99e+03	608	232	38	W87489	S. cerevisiae TIH3 po	2.99e+03
536	5	1.0	108	38	W85919	Amino acid sequence o	2.99e+03	609	232	7	R37805	Human Fc polypeptide	2.99e+03
537	5	1.0	108	22	W07440	Anti-DNA antibody 4b2	2.99e+03	610	234	39	W82694	P. neopolydactyla typ	2.99e+03
538	5	1.0	109	17	R88753	scFv U7.6 VL.	2.99e+03	611	235	39	W89234	Tumour necrosis inhib	2.99e+03
539	5	1.0	110	4	R21287	Murine VL kappa group	2.99e+03	612	241	5	R25245	Chicken GH receptor c	2.99e+03
540	5	1.0	125	5	P20029	Sequence of preprosom	2.99e+03	613	243	39	W82283	Tumour surface protei	2.99e+03
541	5	1.0	128	20	W01404	Generic ob protein.	2.99e+03	614	245	28	W39170	Human PKD1 protein fr	2.99e+03
542	5	1.0	130	38	W73291	R6 monoclonal antibod	2.99e+03	615	246	11	R60522	Anti-TNF diadoby DB32	2.99e+03
543	5	1.0	132	13	R71973	BMP-6.	2.99e+03	616	248	17	R87445	Human glial growth fa	2.99e+03
544	5	1.0	134	7	R38317	Sequence encoded by t	2.99e+03	617	253	38	W85901	Human prion protein (2.99e+03
545	5	1.0	139	39	W84214	Bone morphogenetic pr	2.99e+03	618	254	39	W82285	Natural killer cell p	2.99e+03
546	5	1.0	140	38	W89958	Antigen 2 from cluste	2.99e+03	619	254	38	W85900	Mouse prion protein (2.99e+03
547	5	1.0	140	38	W89957	Antigen 1 from cluste	2.99e+03	620	254	17	R94720	PRRSV VR 2385 ORF-3 p	2.99e+03
548	5	1.0	141	24	W22640	Type II topoisomerase	2.99e+03	621	255	38	W67571	Arabidopsis LEAFY pro	2.99e+03
549	5	1.0	143	9	R49565	Fusion protein JBI22.	2.99e+03	622	255	38	W85903	Sheep prion protein (2.99e+03
550	5	1.0	144	1	P80892	V region of H chain o	2.99e+03	623	256	38	W89870	Antigen from cluster	2.99e+03
551	5	1.0	145	38	W80288	Agrocybe aegerita fru	2.99e+03	624	256	38	W89992	Expressed antigen for	2.99e+03
552	5	1.0	145	3	R15323	IL-2 Chimeric antibod	2.99e+03	625	259	31	W46998	Human telomerase reve	2.99e+03
553	5	1.0	149	38	W82648	Ehrlichia sp. extende	2.99e+03	626	260	38	W89872	Antigen from cluster	2.99e+03
554	5	1.0	151	38	W89809	Protein encoded by c	2.99e+03	627	263	38	W85902	Bovine prion protein	2.99e+03
555	5	1.0	151	38	W81777	Human HE1 protein.	2.99e+03	628	269	14	R87475	Subtilisin N76D/I107V	2.99e+03
556	5	1.0	152	38	W89856	Antigen 2 from cluste	2.99e+03	629	269	39	W82802	Human soluble somatog	2.99e+03
557	5	1.0	154	38	W88690	Secreted protein enco	2.99e+03	630	272	13	R68546	Hepatitis C virus (HC	2.99e+03
558	5	1.0	155	38	W82620	Ehrlichia sp. HGE-23	2.99e+03	631	272	20	W01686	HCV protease clone C3	2.99e+03
559	5	1.0	157	9	R49564	Fusion protein JBI18.	2.99e+03	632	276	38	W88934	Polypeptide fragment	2.99e+03
560	5	1.0	174	39	W84334	GP4 protein of PRRSV	2.99e+03	633	279	7	R34588	Fibronectin domain #2	2.99e+03
561	5	1.0	175	39	W84333	GP4 protein of PRRSV	2.99e+03	634	279	21	W02352	Proteinase K variant	2.99e+03
562	5	1.0	177	39	W68002	Fragment of human sec	2.99e+03	635	280	8	R42055	Insecticidal/Fungicid	2.99e+03
563	5	1.0	178	39	W84328	GP4 protein of PRRSV	2.99e+03	636	281	32	W60776	Amino acids 400 to 68	2.99e+03
564	5	1.0	178	39	W86328	Kidney injury associa	2.99e+03	637	283	2	P70392	Alpha-subunit of Sarc	2.99e+03
565	5	1.0	178	39	W84332	GP4 protein of PRRSV	2.99e+03	638	283	1	P90601	Fibronectin-like poly	2.99e+03
566	5	1.0	178	39	W84331	GP4 protein of PRRSV	2.99e+03	639	290	38	W86337	Kidney injury associa	2.99e+03
567	5	1.0	178	39	W84329	GP4 protein of PRRSV	2.99e+03	640	290	39	W83329	Human mindin-related	2.99e+03
568	5	1.0	178	39	W84324	GP4 protein of PRRSV	2.99e+03	641	291	38	W89923	Antigen from cluster	2.99e+03
569	5	1.0	178	39	W84320	GP4 protein of PRRSV	2.99e+03	642	294	38	W86275	Epichlorohydrin epoxi	2.99e+03
570	5	1.0	178	39	W84325	GP4 protein of PRRSV	2.99e+03	643	294	8	R41513	Mycobacterium fortuit	2.99e+03
571	5	1.0	178	39	W84326	GP4 protein of PRRSV	2.99e+03	644	298	33	W61379	Human junctional adhe	2.99e+03
572	5	1.0	178	39	W84327	GP4 protein of PRRSV	2.99e+03	645	298	38	W85457	Secreted protein enco	2.99e+03
573	5	1.0	178	7	R36537	ICE 20kD subunit.	2.99e+03	646	299	39	W89198	Aspartate-specific cy	2.99e+03
574	5	1.0	183	39	W89689	Murine osteogenic pro	2.99e+03	647	302	5	R28167	Arabidopsis lysozyme/	2.99e+03
575	5	1.0	183	5	R25244	Truncated chicken GH	2.99e+03	648	303	38	W88847	Polypeptide fragment	2.99e+03
576	5	1.0	184	39	W67906	Human secreted protei	2.99e+03	649	303	25	W32380	Mycobacterium tubercu	2.99e+03
577	5	1.0	186	38	W73230	Bovine interferon-bet	2.99e+03	650	310	39	W95197	Human syndecan-1.	2.99e+03
578	5	1.0	186	38	W83336	Mouse ciliary neurotr	2.99e+03	651	310	17	R87000	Human syndecan-1.	2.99e+03
579	5	1.0	186	38	W83901	Bovine interferon-bet	2.99e+03	652	314	28	W34216	Streptomyces cyclase/	2.99e+03
580	5	1.0	186	38	W83900	Bovine interferon-bet	2.99e+03	653	314	19	W02659	G-protein coupled dog	2.99e+03
581	5	1.0	186	38	W83905	Bovine interferon-gam	2.99e+03	654	316	39	W89783	Staphylococcus aureus	2.99e+03
582	5	1.0	186	38	W73229	Feline interferon-bet	2.99e+03	655	319	39	W87701	A human membrane fusi	2.99e+03
583	5	1.0	186	38	W73234	Feline interferon-bet	2.99e+03	656	320	38	W86262	Amino acid sequence o	2.99e+03
584	5	1.0	187	38	W83903	Porcine interferon-be	2.99e+03	657	322	2	R24297	Glycopeptide resistan	2.99e+03
585	5	1.0	187	38	W73232	Porcine interferon-be	2.99e+03	658	326	39	W82311	Human A1 adenosine re	2.99e+03
586	5	1.0	189	38	W83902	Porcine interferon-al	2.99e+03	659	326	16	R87655	Human adenosine recep	2.99e+03
587	5	1.0	189	38	W73231	Porcine interferon-al	2.99e+03	660	328	39	W67852	Human secreted protei	2.99e+03
588	5	1.0	189	5	R25692	Exons XV to XVIII of	2.99e+03	661	328	38	W86336	Kidney injury associa	2.99e+03
589	5	1.0	189	6	R30611	Babesia bovis immunor	2.99e+03	662	329	38	W70801	Amino acid sequence o	2.99e+03
590	5	1.0	191	38	W85579	Human N-methyl-D-aspa	2.99e+03	663	329	18	R91806	Human immunoglobulin	2.99e+03
591	5	1.0	195	7	R34432	Sequence of growth pr	2.99e+03	664	330	5	R29643	pCTD ORF 8.	2.99e+03
592	5	1.0	195	38	W83337	Chicken ciliary neuro	2.99e+03	665	330	3	R15146	EGF-fibronectin fusio	2.99e+03
593	5	1.0	198	3	R12527	Pertussis toxin (delt	2.99e+03	666	331	28	W34215	Streptomyces cyclase/	2.99e+03
594	5	1.0	203	39	W89759	Staphylococcus aureus	2.99e+03	667	331	38	W82660	Maize OTE mature prot	2.99e+03
595	5	1.0	205	22	W20268	H. pylori transporeter	2.99e+03	668	332	3	R15145	EGF-fibronectin fusio	2.99e+03
596	5	1.0	209	38	W82654	Ehrlichia sp. extende	2.99e+03	669	336	39	W86365	Human DNAX toll-like	2.99e+03
597	5	1.0	210	3	R12262	HIV-1 strain OVI open	2.99e+03	670	339	39	W86309	Kidney injury associa	2.99e+03
598	5	1.0	214	39	W82801	Human growth hormone.	2.99e+03	671	341	3	P83200	Sequence encoded by m	2.99e+03
599	5	1.0	215	2	R04026	Q gene product of cDN	2.99e+03	672	342	39	W87644	A protein with water	2.99e+03
600	5	1.0	215	1	P81190	Encoded by gene Y of	2.99e+03	673	345	39	W87715	An active acyltransfe	2.99e+03

674	5	1.0	345 19	W02694	G-protein coupled rat	2.99e+03	747	5	1.0	448 38	W67471	Borage delta-6 desatu	2.99e+03
675	5	1.0	350 2	P70645	Pseudorabies virus gp	2.99e+03	748	5	1.0	448 3	P50132	Sequence of the predo	2.99e+03
676	5	1.0	352 39	W73676	M. tuberculosis antig	2.99e+03	749	5	1.0	450 10	R53263	M. matalensis CCK B r	2.99e+03
677	5	1.0	352 39	W73786	M. tuberculosis antig	2.99e+03	750	5	1.0	454 38	W85046	Gamma subunit of a DN	2.99e+03
678	5	1.0	357 9	R44618	Human retinol binding	2.99e+03	751	5	1.0	454 4	R25155	Fibronectin-spacer-co	2.99e+03
679	5	1.0	360 22	W19703	Melanocortin-3 recept	2.99e+03	752	5	1.0	455 28	W39171	Human PKD1 protein fr	2.99e+03
680	5	1.0	361 39	W73642	Rat adenosine kinase	2.99e+03	753	5	1.0	457 39	W67743	Ovine growth hormone	2.99e+03
681	5	1.0	362 25	W16642	Truncated HSV specifi	2.99e+03	754	5	1.0	461 10	R51002	Sequence of human tum	2.99e+03
682	5	1.0	365 39	W89557	Triticum sp. cysteine	2.99e+03	755	5	1.0	463 21	W14932	Murine anti-porcine V	2.99e+03
683	5	1.0	366 38	W86259	Maize OTE enzyme.	2.99e+03	756	5	1.0	463 8	R41877	Rat glucagon-like pep	2.99e+03
684	5	1.0	367 38	W89051	Polypeptide fragment	2.99e+03	757	5	1.0	464 38	W85045	Gamma subunit of a DN	2.99e+03
685	5	1.0	368 38	W73508	Human ARG-1639 protei	2.99e+03	758	5	1.0	467 22	W05752	Presenilin-1-1 P267S	2.99e+03
686	5	1.0	368 39	W86346	Human FRAZZLED protei	2.99e+03	759	5	1.0	467 4	R22759	Reshaped CD4 antibody	2.99e+03
687	5	1.0	369 39	W89245	Sphingomonas sp. stra	2.99e+03	760	5	1.0	470 39	W83037	Anti-Fas humanised an	2.99e+03
688	5	1.0	369 17	R89276	Rat tyrosine phosphat	2.99e+03	761	5	1.0	470 39	W83036	Anti-Fas humanised an	2.99e+03
689	5	1.0	369 15	R86296	Human tyrosine phosph	2.99e+03	762	5	1.0	472 39	W89559	Triticum sp. cysteine	2.99e+03
690	5	1.0	372 39	W86347	Partial human FRAZZLE	2.99e+03	763	5	1.0	472 39	W86327	Kidney injury associa	2.99e+03
691	5	1.0	372 7	R39590	Trans-acting DNA bind	2.99e+03	764	5	1.0	472 38	W85456	Secreted protein enco	2.99e+03
692	5	1.0	374 39	W88201	Alcohol dehydrogenase	2.99e+03	765	5	1.0	478 21	W14500	Aspergillus oryzae al	2.99e+03
693	5	1.0	374 39	W88205	Alcohol dehydrogenase	2.99e+03	766	5	1.0	480 38	W74814	Human secreted protei	2.99e+03
694	5	1.0	374 39	W88208	Horse liver alcohol d	2.99e+03	767	5	1.0	482 39	W87766	Arabidopsis saccharop	2.99e+03
695	5	1.0	374 39	W88207	Alcohol dehydrogenase	2.99e+03	768	5	1.0	487 4	R09359	Sequence of amylase.	2.99e+03
696	5	1.0	374 39	W88206	Alcohol dehydrogenase	2.99e+03	769	5	1.0	490 39	W89764	Staphylococcus aureus	2.99e+03
697	5	1.0	374 39	W88209	Horse liver alcohol d	2.99e+03	770	5	1.0	493 39	W83125	PrtIIR45 Arginine spe	2.99e+03
698	5	1.0	374 39	W88202	Alcohol dehydrogenase	2.99e+03	771	5	1.0	495 1	P81325	Threonine synthetase.	2.99e+03
699	5	1.0	374 39	W88200	Alcohol dehydrogenase	2.99e+03	772	5	1.0	496 5	R27801	EPSP synthase wild ty	2.99e+03
700	5	1.0	374 39	W88203	Alcohol dehydrogenase	2.99e+03	773	5	1.0	497 13	R72376	Human auxillary cytoc	2.99e+03
701	5	1.0	374 39	W88204	Alcohol dehydrogenase	2.99e+03	774	5	1.0	497 18	R93182	Human cytochrome P450	2.99e+03
702	5	1.0	374 39	W88204	Alcohol dehydrogenase	2.99e+03	775	5	1.0	501 33	W44868	TGF-beta superfamily	2.99e+03
703	5	1.0	374 2	P80277	Sequence encoded by 3	2.99e+03	776	5	1.0	502 38	W81585	CTLA4/p97 fusion prot	2.99e+03
704	5	1.0	374 32	W60774	Amino acids 400 to 77	2.99e+03	777	5	1.0	502 39	W90233	Human sulphamidase pr	2.99e+03
705	5	1.0	376 38	W85054	Beta subunit of DNA P	2.99e+03	778	5	1.0	502 38	W87561	CTLA4-p97 fusion prot	2.99e+03
706	5	1.0	376 25	W35387	Murine metastatic nuc	2.99e+03	779	5	1.0	503 6	R34127	Annexin XI type I fro	2.99e+03
707	5	1.0	377 26	R64143	Gibberellin-20-oxida	2.99e+03	780	5	1.0	504 38	W86245	Mouse ActRIIB4 recept	2.99e+03
708	5	1.0	377 22	W04427	Rat farnesyl transfer	2.99e+03	781	5	1.0	505 28	W39903	HPV16 L1 capsid prote	2.99e+03
709	5	1.0	380 32	W37938	Amino acid sequence o	2.99e+03	782	5	1.0	506 1	P91491	Glycoprotein A antige	2.99e+03
710	5	1.0	380 32	P93309	Promoter of beta-isop	2.99e+03	783	5	1.0	509 39	W95196	Bcl-2 interaction pro	2.99e+03
711	5	1.0	380 1	P90510	Sequence of an epitop	2.99e+03	784	5	1.0	512 38	W86243	Mouse ActRIIB2 recept	2.99e+03
712	5	1.0	387 22	W18663	Fragmented human NF-H	2.99e+03	785	5	1.0	514 39	W83359	Streptococcus pneumon	2.99e+03
713	5	1.0	387 2	P70502	Human protein C inhib	2.99e+03	786	5	1.0	516 39	W85136	A Petunia hybrida EPS	2.99e+03
714	5	1.0	388 38	W73514	Human TGFbetaRII:Fc p	2.99e+03	787	5	1.0	516 39	W85137	Brassica napus EPSPs	2.99e+03
715	5	1.0	388 38	W73513	Rabbit TGFbetaRII:Fc	2.99e+03	788	5	1.0	516 2	P70194	Sequence of chloropla	2.99e+03
716	5	1.0	389 6	R29169	Safflower long-chain	2.99e+03	789	5	1.0	522 9	R48035	Hyaluronic acid synth	2.99e+03
717	5	1.0	393 21	W13949	T284R modified human	2.99e+03	790	5	1.0	527 30	W47207	Homo sapiens tubulin-	2.99e+03
718	5	1.0	394 1	R05282	Amino acid sequence o	2.99e+03	791	5	1.0	528 38	W86244	Mouse ActRIIB3 recept	2.99e+03
719	5	1.0	394 3	P61711	[Ile358] alpha1-antit	2.99e+03	792	5	1.0	528 38	W81457	Canine ICAM-R polypep	2.99e+03
720	5	1.0	394 12	R67363	Alpha-1-antitrypsin m	2.99e+03	793	5	1.0	529 38	W85044	Tau subunit of a DNA	2.99e+03
721	5	1.0	397 38	W86031	Rhizobiaceae phenol d	2.99e+03	794	5	1.0	530 38	W90018	Expressed antigen for	2.99e+03
722	5	1.0	398 3	R15138	Human serotonin 1D re	2.99e+03	795	5	1.0	532 19	R96588	Feline herpesvirus gl	2.99e+03
723	5	1.0	399 38	W86237	Human VEGF-C full len	2.99e+03	796	5	1.0	536 38	W86242	Mouse ActRIIB1 recept	2.99e+03
724	5	1.0	400 38	W88851	Polypeptide fragment	2.99e+03	797	5	1.0	540 18	W03762	Interleukin binding f	2.99e+03
725	5	1.0	402 38	W81966	Human P selectin liga	2.99e+03	798	5	1.0	543 39	W86149	Fat cell differentiat	2.99e+03
726	5	1.0	405 39	W86329	Kidney injury associa	2.99e+03	799	5	1.0	545 33	W50908	Cytophaga drobachiens	2.99e+03
727	5	1.0	409 11	R55790	Cell death hid protei	2.99e+03	800	5	1.0	547 38	W81440	Human intercellular a	2.99e+03
728	5	1.0	412 30	W44190	Herpes simplex virus-	2.99e+03	801	5	1.0	548 22	W20971	H. pylori cytoplasmic	2.99e+03
729	5	1.0	412 38	W81967	Human P selectin liga	2.99e+03	802	5	1.0	549 38	W70898	Acetyl-coenzyme A tra	2.99e+03
730	5	1.0	413 39	W89678	Clostridium chauvoei	2.99e+03	803	5	1.0	551 24	W27150	HMG-CoA reductase deg	2.99e+03
731	5	1.0	418 32	W56709	Amino acid sequence o	2.99e+03	804	5	1.0	555 38	W85133	A desaturase enzyme e	2.99e+03
732	5	1.0	419 38	W86203	Human vascular endoth	2.99e+03	805	5	1.0	555 38	W84154	Human desaturase enzy	2.99e+03
733	5	1.0	420 4	R20200	TGF-alpha-PE40Ab.	2.99e+03	806	5	1.0	562 39	W95512	Amino acid sequence o	2.99e+03
734	5	1.0	420 2	R06450	TGF-alpha-PE40-ab mod	2.99e+03	807	5	1.0	567 16	R88215	Human TGF-beta type I	2.99e+03
735	5	1.0	422 38	W73363	Chicken PAF-AH protei	2.99e+03	808	5	1.0	570 39	W88278	Gracilariaopsis lemane	2.99e+03
736	5	1.0	422 21	W09371	Human neuregulin GGF2	2.99e+03	809	5	1.0	584 23	W18062	Pasteurella haemolyti	2.99e+03
737	5	1.0	422 21	W09372	Human GGF2.	2.99e+03	810	5	1.0	589 39	W85474	Mouse GP88 autocrine	2.99e+03
738	5	1.0	423 17	R91305	Transcription factor-	2.99e+03	811	5	1.0	592 38	W70797	Human interleukin-6R-	2.99e+03
739	5	1.0	424 21	W14765	Human soluble kit lig	2.99e+03	812	5	1.0	593 7	R34669	Ap Serotype 7 65kDa T	2.99e+03
740	5	1.0	429 39	W73631	Human secreted protei	2.99e+03	813	5	1.0	593 39	W85475	Human GP88 autocrine	2.99e+03
741	5	1.0	432 38	W82246	Bacteriophage fd-tet	2.99e+03	814	5	1.0	593 39	W85475	Human GP88 autocrine	2.99e+03
742	5	1.0	438 39	W95500	B. subtilis yabE rela	2.99e+03	815	5	1.0	595 39	R05874	Hsp70 antigen of Myco	2.99e+03
743	5	1.0	439 38	W82645	Ehrlichia sp. extende	2.99e+03	816	5	1.0	595 38	W86003	Anti-5T4 single chain	2.99e+

820	5	1.0	605 21	W12092	Bovine papillomavirus	2.99e+03	893	5	1.0	859 38	W70796	Human gpl30-Fc-His6 a	2.99e+03
821	5	1.0	607 32	W60268	Klebsiella pneumoniae	2.99e+03	894	5	1.0	860 33	W63700	Receptor type tyrosin	2.99e+03
822	5	1.0	608 26	W14645	N. meningitidis IM216	2.99e+03	895	5	1.0	876 16	R80143	B.stearothermophilus	2.99e+03
823	5	1.0	608 38	W84155	Human desaturase enzy	2.99e+03	896	5	1.0	876 16	R80142	B.stearothermophilus	2.99e+03
824	5	1.0	608 38	W85134	A desaturase enzyme e	2.99e+03	897	5	1.0	877 39	W90082	Porcine G-protein reg	2.99e+03
825	5	1.0	608 6	R33646	Rabbit pre-pro serum	2.99e+03	898	5	1.0	878 39	W95397	IBDV VP1 protein.	2.99e+03
826	5	1.0	611 39	W85599	Hexosaminidase enzyme	2.99e+03	899	5	1.0	878 39	W95393	IBDV VP1 protein.	2.99e+03
827	5	1.0	615 39	W95513	Amino acid sequence o	2.99e+03	900	5	1.0	878 38	W81973	Corn pullulanase prot	2.99e+03
828	5	1.0	615 33	W62779	Protein encoded by hu	2.99e+03	901	5	1.0	878 22	W20966	H. pylori flagella-as	2.99e+03
829	5	1.0	620 33	W50907	Alteromonas carragee	2.99e+03	902	5	1.0	880 39	W90088	Human G-protein regul	2.99e+03
830	5	1.0	625 39	W67882	Human secreted protei	2.99e+03	903	5	1.0	884 18	W00160	MEI2 protein kinase P	2.99e+03
831	5	1.0	625 38	W83200	Murine osteoclast dif	2.99e+03	904	5	1.0	885 20	R99740	Human phosphodiestera	2.99e+03
832	5	1.0	626 39	W73533	MEKK3 protein.	2.99e+03	905	5	1.0	889 13	R65159	Potassium ion channel	2.99e+03
833	5	1.0	630 18	R77895	Bacterial transferrin	2.99e+03	906	5	1.0	890 39	W73534	MEKK4 protein.	2.99e+03
834	5	1.0	631 28	W08970	Amino acid sequence o	2.99e+03	907	5	1.0	895 28	W34258	Rat ob receptor isofo	2.99e+03
835	5	1.0	636 12	R68576	Rat MDF encoded by cl	2.99e+03	908	5	1.0	908 38	W85587	Human N-methyl-D-aspa	2.99e+03
836	5	1.0	638 1	R05589	Acetohydroxy acid syn	2.99e+03	909	5	1.0	924 38	W73511	Human neuronal ICAM-4	2.99e+03
837	5	1.0	640 39	W87762	Maize lysine ketoglut	2.99e+03	910	5	1.0	929 38	W87501	N-methyl-D-aspartate	2.99e+03
838	5	1.0	655 10	R53962	Hepatocyte growth fac	2.99e+03	911	5	1.0	933 39	W89200	Human phosphatidylcho	2.99e+03
839	5	1.0	661 39	W87556	B cell surface protei	2.99e+03	912	5	1.0	936 39	W89801	Staphylococcus aureus	2.99e+03
840	5	1.0	663 24	W29150	Dual-specific murine	2.99e+03	913	5	1.0	945 39	W73624	Human secreted protei	2.99e+03
841	5	1.0	675 20	W06449	Brassica napus FCA po	2.99e+03	914	5	1.0	948 32	W48781	Thyroid peroxidase.	2.99e+03
842	5	1.0	686 13	R66267	Hepatitis C virus (HC	2.99e+03	915	5	1.0	950 39	W83319	Mouse SMAD interactin	2.99e+03
843	5	1.0	686 20	W01693	Hepatitis C virus NS3	2.99e+03	916	5	1.0	951 38	W70798	Human gpl30-C-gamma-1	2.99e+03
844	5	1.0	688 21	W06316	TRP-1 protein.	2.99e+03	917	5	1.0	955 8	R42241	Human p50 protein DNA	2.99e+03
845	5	1.0	693 14	R85146	Lactoferrin protein.	2.99e+03	918	5	1.0	955 8	R42246	Human p50 protein DNA	2.99e+03
846	5	1.0	698 39	W83493	4D5 Fab molecule expr	2.99e+03	919	5	1.0	957 22	W20841	H. pylori transmembra	2.99e+03
847	5	1.0	703 38	W86023	Porcine lactoferrin (2.99e+03	920	5	1.0	965 31	W56567	Toxin TccA, encoded b	2.99e+03
848	5	1.0	708 38	W86022	Bovine lactoferrin (h	2.99e+03	921	5	1.0	967 10	R51695	PEPC.	2.99e+03
849	5	1.0	709 1	P91934	Bi antigen.	2.99e+03	922	5	1.0	970 39	W89409	Maize phosphoenolpyru	2.99e+03
850	5	1.0	711 38	W86021	Human lactoferrin (hL	2.99e+03	923	5	1.0	976 38	W85581	Human N-methyl-D-aspa	2.99e+03
851	5	1.0	713 17	R85878	WD-40 domain-contg. T	2.99e+03	924	5	1.0	979 39	W83068	TutC protein of tolu	2.99e+03
852	5	1.0	718 3	R14308	N.clavipes dragline s	2.99e+03	925	5	1.0	985 8	R42995	Glycosyltransferase.	2.99e+03
853	5	1.0	722 18	R98227	Rat neuronal protein	2.99e+03	926	5	1.0	996 38	W85033	IkappaB kinase alpha	2.99e+03
854	5	1.0	728 39	W88531	Human hepatocyte grow	2.99e+03	927	5	1.0	997 38	W85032	Green fluorescent pro	2.99e+03
855	5	1.0	728 39	W88530	Human hepatocyte grow	2.99e+03	928	5	1.0	997 38	W85589	Human N-methyl-D-aspa	2.99e+03
856	5	1.0	728 39	W88529	Human hepatocyte grow	2.99e+03	929	5	1.0	1012 39	W95395	IBDV VP2-VP3-VP4 poly	2.99e+03
857	5	1.0	728 39	W88532	Human hepatocyte grow	2.99e+03	930	5	1.0	1022 39	W87761	Maize lysine ketoglut	2.99e+03
858	5	1.0	728 8	R52945	Human hepatocyte grow	2.99e+03	931	5	1.0	1039 38	W85014	Stat5-green flouresce	2.99e+03
859	5	1.0	728 8	R40862	Competative inhibitor	2.99e+03	932	5	1.0	1045 38	W85050	Stat5-green flouresce	2.99e+03
860	5	1.0	730 22	W13669	C-proteinase encoded	2.99e+03	933	5	1.0	1045 39	W86354	Human DNAX toll-like	2.99e+03
861	5	1.0	742 1	R05235	Amino acid sequence o	2.99e+03	934	5	1.0	1064 39	W87759	Arabidopsis lysine ke	2.99e+03
862	5	1.0	745 39	W82498	Human IKK-alpha prote	2.99e+03	935	5	1.0	1066 39	W88255	Morchella costata alp	2.99e+03
863	5	1.0	745 39	W86164	Human IKK-alpha polyp	2.99e+03	936	5	1.0	1070 39	W88256	Morchella vulgaris al	2.99e+03
864	5	1.0	745 38	W81562	IKB kinase (IKK)-alph	2.99e+03	937	5	1.0	1086 38	W81839	Human LFA-1 alpha cha	2.99e+03
865	5	1.0	745 38	W81565	IKB kinase (IKK)-alph	2.99e+03	938	5	1.0	1092 39	W88257	Gracilariaopsis lemne	2.99e+03
866	5	1.0	746 38	W85135	A desaturase enzyme e	2.99e+03	939	5	1.0	1099 20	W05177	Lats large tumour sup	2.99e+03
867	5	1.0	746 38	W84156	Human desaturase enzy	2.99e+03	940	5	1.0	1118 38	W82395	Human UBP protein #1.	2.99e+03
868	5	1.0	752 38	W82646	Ehrlichia sp. extende	2.99e+03	941	5	1.0	1125 30	W52288	Rattus norvegicus cdo	2.99e+03
869	5	1.0	753 39	W95514	Amino acid sequence o	2.99e+03	942	5	1.0	1130 20	W05178	H-lats large tumour s	2.99e+03
870	5	1.0	764 18	R91809	Human IgG CH1-hinge-C	2.99e+03	943	5	1.0	1132 31	W46957	Human telomerase reve	2.99e+03
871	5	1.0	769 32	W60795	Deletion mutant delta	2.99e+03	944	5	1.0	1134 29	W46859	Bacillus thuringiens	2.99e+03
872	5	1.0	770 23	W19485	APP770 mutant A-beta-	2.99e+03	945	5	1.0	1151 38	W73344	Rat alphad protein se	2.99e+03
873	5	1.0	774 26	W35390	Flavobacterium odorat	2.99e+03	946	5	1.0	1155 38	W73346	Mouse alphad protein se	2.99e+03
874	5	1.0	774 17	R93984	Cephalosporin C acyla	2.99e+03	947	5	1.0	1161 38	W73345	Rat alphad protein se	2.99e+03
875	5	1.0	774 15	R86374	Cephalosporin C acyla	2.99e+03	948	5	1.0	1161 38	W73347	Mouse alphad protein	2.99e+03
876	5	1.0	774 15	R86373	Cephalosporin C acyla	2.99e+03	949	5	1.0	1171 38	W85037	Green fluorescent pro	2.99e+03
877	5	1.0	777 39	W85601	Hexosaminidase enzyme	2.99e+03	950	5	1.0	1181 38	W85038	NFAT1-green flouresce	2.99e+03
878	5	1.0	777 39	W54079	Homo sapiens BARD1 se	2.99e+03	951	5	1.0	1184 20	W10676	Guanine thymine bindi	2.99e+03
879	5	1.0	784 39	W87760	Soybean lysine ketogl	2.99e+03	952	5	1.0	1189 2	R08399	Human Serum Albumin a	2.99e+03
880	5	1.0	789 29	W46861	Bacillus thuringiens	2.99e+03	953	5	1.0	1214 38	W87509	Human N-methyl-D-aspa	2.99e+03
881	5	1.0	798 38	W85025	Smad4-green flouresce	2.99e+03	954	5	1.0	1219 38	W87507	Human N-methyl-D-aspa	2.99e+03
882	5	1.0	799 39	W86352	Human DNAX toll-like	2.99e+03	955	5	1.0	1229 39	W87633	CryET5 protein sequen	2.99e+03
883	5	1.0	806 16	R85301	Tomato pathogen resis	2.99e+03	956	5	1.0	1231 38	W87506	Human N-methyl-D-aspa	2.99e+03
884	5	1.0	806 30	W53103	Anabaena sucrose synt	2.99e+03	957	5	1.0	1236 38	W85574	Human N-methyl-D-aspa	2.99e+03
885	5	1.0	806 38	W85013	Smad4-green flouresce	2.99e+03	958	5	1.0	1239 38	W87508	Human N-methyl-D-aspa	2.99e+03
886	5	1.0	809 28	W19521	B. cereus VIP3A(a) sy	2.99e+03	959	5	1.0	1244 38	W87505	Human N-methyl-D-aspa	2.99e+03
887	5	1.0	810 39	W73471	Mouse interleukin-4 r	2.99e+03	960	5	1.0	1247 39	W73535	MEKK5 protein.	2.99e+03
888	5	1.0	834 22	W13506	Human signal mediator	2.99e+03	961	5	1.0	1274 39	W89253	Human ALP.	2.99e+03
889	5	1.0	837 39	W86361	Human DNAX toll-like	2.99e+03	962	5	1.0	1323 10	R55248	Human N-methyl-D-aspa	2.99e+03
890	5	1.0	841 28	W46397	Amino acid sequence o	2.99e+03	963	5	1.0	1336 38	W87511	Tissue-binding hybrid	2.99e+03
891	5	1.0	847 39	W85603	Hexosaminidase enzyme	2.99e+03	964	5	1.0	1336 11	R60019	Aldehyde oxidase prot	2.99e+03
892	5	1.0	849 38	W82397	Human UBP protein #3.	2.99e+03	965	5	1.0	1358 33	W57837		


```
966 5 1.0 1363 2 R11061 Bovine Coronavirus E2 2.99e+03
967 5 1.0 1451 38 W83310 LRP5 protein from iso 2.99e+03
968 5 1.0 1454 9 R42474 FECV/FIPV chimeric sp 2.99e+03
969 5 1.0 1479 3 R13231 CFTR delta I507. 2.99e+03
970 5 1.0 1480 3 R13232 CFTR G85E. 2.99e+03
971 5 1.0 1484 38 W87510 Human N-methyl-D-aspa 2.99e+03
972 5 1.0 1493 5 R27640 Human calcium channel 2.99e+03
973 5 1.0 1561 18 W02098 S. mutans antigen I/I 2.99e+03
974 5 1.0 1565 31 W56568 Toxin TccB, encoded b 2.99e+03
975 5 1.0 1591 38 W83308 Mature LRP5 protein. 2.99e+03
976 5 1.0 1614 38 W83312 Mouse Lrp5 protein. 2.99e+03
977 5 1.0 1615 38 W83309 LRP5 protein from the 2.99e+03
978 5 1.0 1639 38 W83311 LRP5 isoform 3 protei 2.99e+03
979 5 1.0 1706 24 W24786 Pitr antigenic protei 2.99e+03
980 5 1.0 1763 39 W83431 Rat Munc13-1. 2.99e+03
981 5 1.0 1852 22 W10011 Protein encoded by mu 2.99e+03
982 5 1.0 1852 19 R99448 BRCA1 allele #77 tran 2.99e+03
983 5 1.0 1863 18 R81539 BRCA1 mutant from PM1 2.99e+03
984 5 1.0 1863 18 R81514 BRCA1 mutant from sam 2.99e+03
985 5 1.0 1863 18 R81511 BRCA1 mutant from sam 2.99e+03
986 5 1.0 1938 39 W73553 Full length Pullulana 2.99e+03
987 5 1.0 2100 39 W89579 Calcium permeable vol 2.99e+03
988 5 1.0 2104 39 W89578 Calcium permeable vol 2.99e+03
989 5 1.0 2105 39 W89577 Calcium permeable vol 2.99e+03
990 5 1.0 2150 2 R05127 Complete human rhinov 2.99e+03
991 5 1.0 2209 5 P20037 Sequence encoded by a 2.99e+03
992 5 1.0 2324 16 R92778 Human fibronectin. 2.99e+03
993 5 1.0 2435 5 R25135 HCV polypeptide 1. 2.99e+03
994 5 1.0 2639 39 W73476 Grapevine leafroll vi 2.99e+03
995 5 1.0 2766 22 W19651 ATM mutant 8283delTC. 2.99e+03
996 5 1.0 2799 39 W81867 Human tumour suppress 2.99e+03
997 5 1.0 2938 11 R59923 GAP protein Iral. 2.99e+03
998 5 1.0 3033 6 R33214 NANBH virus strain HC 2.99e+03
999 5 1.0 3085 22 W19701 ATM mutant G9170C. 2.99e+03
1000 5 1.0 3898 2 R06996 Protein characteristi 2.99e+03
```

ALIGNMENTS

```
RESULT 1
ID R47466 standard; Protein; 503 AA.
AC R47466;
DT 23-JUN-1994 (first entry)
DE Ced-3.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW Protein synthesis; null phenotype; calcium-binding domain.
KW Caenorhabditis elegans.
PN W09325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.
DR N-PSDB; Q54666.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PT Claim 6; Fig 4; 127pp; English.
PS This sequence is encoded by the C. elegans ced-3 gene. A 2.8 kb mRNA
CC was identified as the ced-3 transcript and was most abundant in
CC embryos, but was also detected in larvae and young adults, suggesting
CC that ced-3 is expressed not only in cells undergoing cell death. The
CC four largest introns as well as sequences 5' of the start codon
CC contain repetitive elements, some of which have been characterised
CC in non-coding regions of other C. elegans genes, such as fem-1, lin-12
CC and myoD. The Ced-3 protein is 503 amino acids in length. Ced-3 is
CC highly hydrophilic with no significant hydrophobic region that might
CC be a transmembrane region. One region of Ced-3 is very rich in serine.
CC It is thought that this region is involved in protein-protein inter-
CC actions, similar to acid blobs in transcription factors. Of the
```

```
CC mutations which occur within the ced-3 gene, eight of the mutations
CC are missense mutations, two are nonsense mutations and two are putative
CC splicing mutations (see also R53279-88). These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 503 AA;
```

```
Query Match 100.0%; Score 503; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Db 1 mmrgdrslrnimmfsshlkvdeilevliakqvlnsdngdminscgtvrekrrreivka 60
QY |||||
1 MMQRDRSLRNIMMFSSHLKVDEILEVLIKQVLNSDNGDMINSCGTVREKRREIVKA 60

Db 61 vqrgdvafdafydlrstgheglavleplarsvdsnavefecmpspashrrsalspa 120
QY |||||
61 VQRGDAVDAFYDALRSTGHEGLAEVLEPLARSDVSNVAFECMPSPASHRRSRALSPA 120

Db 121 gytsptrvhrdsvssvstsyqdiysrarsrsralhssdrhnyssppvnaafpsqps 180
QY |||||
121 GYTSPTRVHRDSVSSVSTSYQDIYSRARSRSRALHSSDRHNYSSPPVNAFPSPSS 180

Db 181 ansftgcsslgysssrnsrfskasgptqyifheedmnfdaptisrvfdektmyrnfs 240
QY |||||
181 ANSFTGCSSLGYSSSRNSRFSKASGPTQYIFHEEDMNFVDAPTISRVDKTMRYNFSS 240

Db 241 prgmclinnnehfeqmptrngtkadkdnltlfrcmgytvickdnltgrgmlltirdfak 300
QY |||||
241 PRGMCLINNNEHFEQMPTRNGTKADKDNLTNLFRCMGYTVICKDNLTGRGMLLTIRDFAK 300

Db 301 heshgdsailvilshgeenviigvddipistheiydlalnaaprlankpkivfvqacrg 360
QY |||||
301 HESHGDSAILVILSHGEENVIGVDDIPISTHEIYDLLNAANAPRLANKPKIVFVQACRG 360

Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnflgcvrqpqvvrkkpsqadiliry 420
QY |||||
361 ERRDNGFPVLDSDVGPAPFLRRGWDNRDGPLFNFLGCVRPQVQVVRKKPSQADILIRYA 420

Db 421 ttaqvsvwrnsargswfiqavcevfsthakdmvveltevnkvacgftsqgsnilkq 480
QY |||||
421 TTAQVSVWRNSARGSWFIQAVCEVFSTHAKDMDVVELLTVNKKVACGFTSQGSNILKQ 480

Db 481 mpemtsrlkkfyfwpearnsav 503
QY |||||
481 MPEMTSRLKKFYFWPEARNSAV 503
```

```
RESULT 2
ID R98754 standard; Protein; 503 AA.
AC R98754;
DT 17-DEC-1996 (first entry)
DE Nematode Ced-3 cell death protein.
KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;
KW cell death; apoptosis; neural degeneration; inflammation;
KW antiinflammatory.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT region 78..205
FT /label= Serine-rich_region
FT cleavage_site 131
FT /note= "autocleavage site"
FT cleavage_site 371
FT /note= "autocleavage site"
FT active_site 358
FT /note= "active site Cysteine residue"
PN W09625946-A1.
PD 29-AUG-1996.
PF 23-FEB-1996; U02473.
```



```
PR 24-FEB-1995; US-394189.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 96-425082/42.
DR N-PSDB; T38196.
PT Ced-3 and human interleukin 1-beta convertase genes and proteins
PT useful to treat inflammation and diseases characterised by cell
PT death
CC Claim 18; Fig 3; 139pp; English.
CC The Caenorhabditis elegans Ced-3 protein (R98754) is encoded
CC by the ced-3 cell death gene (T38196). It is structurally similar
CC to human interleukin-1 beta converting enzyme (ICE) (R98755); the
CC 2 proteins have 28% amino acid identity (higher in the C-terminal
CC regions critical for activity), suggesting that Ced-3 may be a
CC cysteine protease like ICE and that ICE may be a human equivalent
CC of the nematode cell death protein. Ced-3 proteins can therefore
CC be used in prodn. of novel drugs for enhancing or inhibiting the
CC activity of ICE for the treatment of inflammatory diseases and/or
CC diseases caused by cell death. Novel inhibitors of ced-3 activity
CC include portions of Ced-3. Ced-3 with a C360A substn. (see also
CC R98758) can prevent programmed cell death in C. elegans.
SQ Sequence 503 AA;

Query Match 100.0%; Score 503; DB 19; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mmrqdrslrllernimfshhlkvdeilevliakqvlnsdngminscgtvrekrrreivka 60
QY 1 MMRQDRSLRLERNIMFSSHLKVDEILEVLIKQVLNSDNGMINSCGTVREKRREIVKA 60

Db 61 vqrrgdvafdafydlrstgheglaeleplarsvdsnavfecmpspashrrsalspa 120
QY 61 VQRRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDNSNAVEFECMPSPASHRRSRALSPA 120

Db 121 gytsptrvhrdsvssvssftsytqdiysrarsrsralhssdrhnyssppvnafpsqss 180
QY 121 GYTSPTRVHRDSVSSVSSFTSYQDIYSRARSRSRALHSSDRHNYSSPPVNAFPSPSS 180

Db 181 anssftgcsslgysssrnrsfksagptqyifheedmnfvdpatisrvfdektmyrnfs 240
QY 181 ANSSFTGCSSLGYSSSRNRSFSKASGPTQYIFHEEDMNFVDAPTISRVEKTMRYNFS 240

Db 241 prgmclliinnehefemqptnrgtkadkdnltlncfcmgytvickdnltrgrmlltirdfak 300
QY 241 PRGMCLLIINNEHEFEMQPTNRTGKTADKDNLTNLCFRCMGYTVICKDNLTRGRMLLTIRDFAK 300

Db 301 heshgdsailvilshgeenviigvddipistheiydlldnaaprlankpkivfvqacrg 360
QY 301 HESHGDSAILVILSHGEENVIIIGVDDIPISTHEIYDILLNAANAPRLANKPKIVFVQACRG 360

Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnflgcvrpgvqgvwrkksqadiliry 420
QY 361 ERRDNGFPVLDSVDGVPFAFLRRGWDNRDGPLFNFLGCVRPQVQVWRKKPSQADILIRYA 420

Db 421 ttaqvswrnsargswfiqavcevfsthakdmdvveltevnkvacgftsqgsnilkq 480
QY 421 TTAQVSWRNSARGSWFIQAVCEVFSTHAKDMDVVELLTVENKVKACGFTSQGSNILKQ 480

Db 481 mpemtsrllkkkfyfwpearnasv 503
QY 481 MPEMTSRLKKKFFWPEARNASV 503

RESULT 3
ID R53288 standard; Protein; 503 AA.
AC R53288;
DT 23-JUN-1994 (first entry)
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain.
```

```
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT misc_difference 486
FT /label= S486F
FT /note= "From allele n1163"
PN WO9325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.
DR N-PSDB; Q64745.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PT Claim 13; Fig 4; 127pp; English.
PS The sequences given in R53279-88 are encoded by mutant versions of the
CC C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
CC transcript and was most abundant in embryos, but was also detected in
CC larvae and young adults, suggesting that ced-3 is expressed not only
CC in cells undergoing cell death. The four largest introns as well as
CC sequences 5' of the start codon contain repetitive elements, some of
CC which have been characterised in non-coding regions of other C. elegans
CC genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
CC acids in length. Ced-3 is highly hydrophilic with no significant
CC hydrophobic region that might be a transmembrane region. One region
CC of Ced-3 is very rich in serine. It is thought that this region is
CC involved in protein-protein interactions, similar to acid blobs in
CC transcription factors. Of the mutations which occur within the ced-3
CC gene, eight of the mutations are missense mutations. Two are nonsense
CC mutations and two are putative splicing mutations. These mutations
CC establish the null phenotype of the ced-3 gene, confirming that ced-3,
CC like ced-4, function is not essential for viability. The ced-3 and
CC ced-4 gene products may be used to develop agents for treating
CC conditions characterised by cell deaths, such as myocardial infarction,
CC stroke, degenerative disease, traumatic brain injury, hypoxia,
CC pathogenic infection, aging or hair loss.
SQ Sequence 503 AA;
```

```
Query Match 96.4%; Score 485; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mmrqdrslrllernimfshhlkvdeilevliakqvlnsdngminscgtvrekrrreivka 60
QY 1 MMRQDRSLRLERNIMFSSHLKVDEILEVLIKQVLNSDNGMINSCGTVREKRREIVKA 60

Db 61 vqrrgdvafdafydlrstgheglaeleplarsvdsnavfecmpspashrrsalspa 120
QY 61 VQRRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDNSNAVEFECMPSPASHRRSRALSPA 120

Db 121 gytsptrvhrdsvssvssftsytqdiysrarsrsralhssdrhnyssppvnafpsqss 180
QY 121 GYTSPTRVHRDSVSSVSSFTSYQDIYSRARSRSRALHSSDRHNYSSPPVNAFPSPSS 180

Db 181 anssftgcsslgysssrnrsfksagptqyifheedmnfvdpatisrvfdektmyrnfs 240
QY 181 ANSSFTGCSSLGYSSSRNRSFSKASGPTQYIFHEEDMNFVDAPTISRVEKTMRYNFS 240

Db 241 prgmclliinnehefemqptnrgtkadkdnltlncfcmgytvickdnltrgrmlltirdfak 300
QY 241 PRGMCLLIINNEHEFEMQPTNRTGKTADKDNLTNLCFRCMGYTVICKDNLTRGRMLLTIRDFAK 300

Db 301 heshgdsailvilshgeenviigvddipistheiydlldnaaprlankpkivfvqacrg 360
QY 301 HESHGDSAILVILSHGEENVIIIGVDDIPISTHEIYDILLNAANAPRLANKPKIVFVQACRG 360

Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnflgcvrpgvqgvwrkksqadiliry 420
QY 361 ERRDNGFPVLDSVDGVPFAFLRRGWDNRDGPLFNFLGCVRPQVQVWRKKPSQADILIRYA 420

Db 421 ttaqvswrnsargswfiqavcevfsthakdmdvveltevnkvacgftsqgsnilkq 480
```



```

QY 421 TTAQYVSWRNSARGSWFIQAVCEVFSTHAKDMVDVVELLTVNKKVACGFTSQGSNILKQ 480
Db 481 mpemt 485
QY 481 MPEMT 485

RESULT 4
ID R53287 standard; Protein; 503 AA.
AC R53287;
DT 23-JUN-1994 (first entry)
DE Ced-3 (E483K).
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain.
OS Caenorhabditis elegans.
FH Key
FT Location/Qualifiers
FT misc_difference 483
FT /label= E483K
FT /note= "From allele n2426"

W09325685-A.
23-DEC-1993.
14-JUN-1993; U05701.
12-JUN-1992; US-897788.
20-NOV-1992; US-979638.
(MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.
NP N-PSDB; Q64744.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PT Claim 13; Fig 4; 127pp; English.
PS The sequences given in R53279-88 are encoded by mutant versions of the
CC C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
CC transcript and was most abundant in embryos, but was also detected in
CC larvae and young adults, suggesting that ced-3 is expressed not only
CC in cells undergoing cell death. The four largest introns as well as
CC sequences 5' of the start codon contain repetitive elements, some of
CC which have been characterised in non-coding regions of other C. elegans
CC genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
CC acids in length. Ced-3 is highly hydrophilic with no significant
CC hydrophobic region that might be a transmembrane region. One region
CC of Ced-3 is very rich in serine. It is thought that this region is
CC involved in protein-protein interactions, similar to acid blobs in
CC transcription factors. Of the mutations which occur within the ced-3
CC gene, eight of the mutations are missense mutations, two are nonsense
CC mutations and two are putative splicing mutations. These mutations
CC establish the null phenotype of the ced-3 gene, confirming that ced-3,
CC like ced-4, function is not essential for viability. The ced-3 and
CC ced-4 gene products may be used to develop agents for treating
CC conditions characterised by cell deaths, such as myocardial infarction,
CC stroke, degenerative disease, traumatic brain injury, hypoxia,
CC pathogenic infection, aging or hair loss.
SQ Sequence 503 AA;

Query Match 95.8%; Score 482; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 482; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mmrqdrslrnlmfmfshlkvdeilevliakqvlnsdngdminscgtvrekrrvika 60
QY 1 MMRQDRSLRLERNIMFMFSSHLKVDLEILEVLIKQVLNSDNGDMINSCGTVREKRRVKA 60
Db 61 vqrrgdvafdafydlarstgheglavleplarsvdsnavefecmpspashrralspa 120
QY 61 VQRRGDVAFDAFYDLARSTGHEGLAEVLEPLARSVDSDNAVEFECMPSPASHRRALS SPA 120
Db 121 gytsptrhrdsvssvssftsytqdiysrarsrslhssdrhnyssppvnafpsqps 180
QY 121 GYTSPTVRHRDSVSSVSSFTSYQDIYSRARSRSRALHSSDRHNYSSPPVNAFPSPSS 180
```

```

Db 181 anssftgcsslgysssrnsrfskasgptqyifheednmfvdaptisrvfdektmyrnfs 240
QY 181 ANSFTGCSSLGYSSSRNRSFSKASGPTQYIFHEEDMNFVDAPTISRVEDEKTMYNFSS 240

Db 241 prgmcllinnehfeqmptrngtkadkdnltlnlfrcmgytvtickdnltgrgmlltirdfak 300
QY 241 PRGMCLIIINNEHFEOQMPTRNGTKADKDNLTNLFRCMGYTVICKDNLTGRGMLLTIRDFAK 300

Db 301 heshgdsailvlsbgeenviigvddipistheiydlinaanaprlankpkivfvqacrg 360
QY 301 HESHGDSAILVLSHGEENVIIIGVDDIPISTHEIYDLLNAANAPRLANKPKIVFVQACRG 360

Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnlglcvrpgqvwrkkpsqadiliry 420
QY 361 ERRDNGFPVLDSVDGVPFAFLRRGWDNRDGLFNFLGCVRPQVQVWRKKPSQADILIRYA 420

Db 421 ttaqvswrnsargswfiqavcevfsthakdmvdvveltevnkvcgftsqgsnilkq 480
QY 421 TTAQYVSWRNSARGSWFIQAVCEVFSTHAKDMVDVVELLTVNKKVACGFTSQGSNILKQ 480

Db 481 mp 482
QY 481 MP 482

RESULT 5
ID R53279 standard; Protein; 503 AA.
AC R53279;
DT 23-JUN-1994 (first entry)
DE Ced-3 (L27F).
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain.
OS Caenorhabditis elegans.
FH Key
FT Location/Qualifiers
FT misc_difference 27
FT /label= L27F
FT /note= "From allele n1040"

W09325685-A.
23-DEC-1993.
14-JUN-1993; U05701.
12-JUN-1992; US-897788.
20-NOV-1992; US-979638.
(MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.
DR N-PSDB; Q64735.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PT Claim 13; Fig 4; 127pp; English.
PS The sequences given in R53279-88 are encoded by mutant versions of the
CC C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
CC transcript and was most abundant in embryos, but was also detected in
CC larvae and young adults, suggesting that ced-3 is expressed not only
CC in cells undergoing cell death. The four largest introns as well as
CC sequences 5' of the start codon contain repetitive elements, some of
CC which have been characterised in non-coding regions of other C. elegans
CC genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
CC acids in length. Ced-3 is highly hydrophilic with no significant
CC hydrophobic region that might be a transmembrane region. One region
CC of Ced-3 is very rich in serine. It is thought that this region is
CC involved in protein-protein interactions, similar to acid blobs in
CC transcription factors. Of the mutations which occur within the ced-3
CC gene, eight of the mutations are missense mutations, two are nonsense
CC mutations and two are putative splicing mutations. These mutations
CC establish the null phenotype of the ced-3 gene, confirming that ced-3,
CC like ced-4, function is not essential for viability. The ced-3 and
CC ced-4 gene products may be used to develop agents for treating
CC conditions characterised by cell deaths, such as myocardial infarction,
CC stroke, degenerative disease, traumatic brain injury, hypoxia,
CC pathogenic infection, aging or hair loss.
SQ Sequence 503 AA;
```


	Query Match	9.7%;	Score 350;	DB 11;	Length 277;
	Best Local Similarity	34.9%;	Pred. No. 8.07e-49;		
	Matches	97;	Conservative	59;	Mismatches 97; Indels 25; Gaps 17
Db	10 SREVLDPAEQYKMDHKRRGTALIFNHERFEFWHLALPERRGTNADRDNPTRRSELGFEVK	69			
QY	226 SR-VFDEKTMYNRFSSPRGMCLINNEHF--E-QMPTRNGTKADKDNLTFNFCRMGYTVI	281			
Db	70 CFNDLRAEELLKIHEVST-SSHVDADCFLCVFLSHGEGNHYYAYDAK-IETQLTLGLEK	127			
QY	282 CKDNLTRGMILLTRIDFAKHESHGDSA-IL-VILSHGEENVIIIGVDDIPISITHEIYDLLN	339			
Db	128 GDKCQSLVGPKIFIIOACRGSQHNV--P-LVPLDVV-D-HQT-DKLDD--N-VTQY-	175			
QY	340 AANAPRLANKPKIVEVOACRGERDRNGFPVLDSDYGVPAPFLRRGWDRNRDGPLENFGLCYR	399			
Db	176 -DAASYV-TLPAGADFLMCYSVAEGYISHRETIVNGSWYIQDLSEMALARHGSSLEFTELLT	233			
QY	400 PQYQQYWRKKPSQADILIRYATTAQYVSWRNSARGSWFIQAVCEVFSTHAKDMVVELLT	459			
Db	234 LVNRKYSQRRVDFCKDPGAIGKKQVPCFASMLTKKLHF	271			
QY	460 EVNKKVA---CGFQTSQSGSNILKQMPEMTSRLLKKFYF	494			
RESULT	6	PRELIMINARY;	PRT;	303	AA.
ID	088550				
AC	088550;				

RESULT	7			
ID	042284	PRELIMINARY;	PRT;	283 AA.
AC	042284;			
DT	01-JAN-1998 (TREMBLREL. 05, CREATED)			
DT	01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	CASPASE-1.			
OS	GALLUS GALLUS (CHICKEN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;			
OC	NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-WHITE LEGHORN;			
RA	JOHNSON A.L., BRIDGHAM J.T., MUNKS M.W.;			
RL	SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.			
DR	EMBL; AF031351; G2642241; -			
DR	PFAM; PF00655; ICE_p10; 1.			
DR	PFAM; PF00656; ICE_p20; 1.			
SQ	SEQUENCE 283 AA; 32622 MW; 7A72DE81 CRC32;			
Query Match	8.3%;	Score 298;	DB 13;	Length 283;
Best Local Similarity	31.98;	Pred. No. 1.35e-37;		
Matches	82;	Conservative 56;	Mismatches 93;	Indels 26;
			Gaps 19;	
DB	45 LLICNTKFKHLRQRDGAENVDKEMTKLEGLGVNEVEDKEDKTSQEMTTVMKKFADHKDHL 104			
QY	246 LIINNEHEQMPTRNGTKADKDNLTNLFRCMGTYVICKDNLTRGRMLITIRDFAKHESH- 304			
DB	105 TSDSTFLVEMSHGWMSTGICGCTKSNCTDILSFNTIYENFNKKHCRALMGKPKVVIQCCR 164			


```
Db 61 QRRGVAFDAFYDALRSTGHEGLAEVLEPLARSVDNAVEFECMPSPASHRSRALSPAG 120
QY 62 QRRGVAFDAFYDALRSTGHEGLAEVLEPLARSVDNAVEFECMPSPASHRSRALSPAG 121
Db 121 YTSPTRVHRDSSVSSSFTSYQDIYSRARSRSRSLHSSDRHNSSPPVNAFPSSQSSA 180
QY 122 YTSPTRVHRDSSVSSSFTSYQDIYSRARSRSRSLHSSDRHNSSPPVNAFPSSQSSA 181
Db 181 NSSFTGCGSSLGSSSSSRNRSEKASGPTQYIFHEEDMNFVDAPTISRVEDEKMYRNSSP 240
QY 182 NSSFTGCGSSLGSSSSSRNRSEKASGPTQYIFHEEDMNFVDAPTISRVEDEKMYRNSSP 241
Db 241 RGMCLINNEHEFQMPTRNGTKADKDNLTNLFRCMGYIVICKDNLTRGMLLTIRDFAKH 300
QY 242 RGMCLINNEHEFQMPTRNGTKADKDNLTNLFRCMGYIVICKDNLTRGMLLTIRDFAKH 301
Db 301 ESHGDSAILVILSHGEENVIGVDDIPISTHEIYDLNANAPRLANKPKIYVQACRGE 360
QY 302 ESHGDSAILVILSHGEENVIGVDDIPISTHEIYDLNANAPRLANKPKIYVQACRGE 361
Db 361 RRDNGFPVLDSDVGPAPFLRGMWNRDGPLFNLGCVRPQOVWRKKPSQADILIRYAT 420
QY 362 RRDNGFPVLDSDVGPAPFLRGMWNRDGPLFNLGCVRPQOVWRKKPSQADILIRYAT 421
Db 421 TAQYVSWRNSARGSWFIQAVCEVSTHAKMDVVELLTVENKKVAYDIPPAQ 472
QY 422 TAQYVSWRNSARGSWFIQAVCEVSTHAKMDVVELLTVENKKVACGFTSQ 473
```

```
RESULT 2
ID 093415 PRELIMINARY; PRT; 304 AA.
AC 093415;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CASPASE-6.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA JOHNSON A.L., BRIDGHAM J.T.;
RT "Characterization of the chicken caspase-6 cDNA."
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF082329; G3435272; -.
SQ SEQUENCE 304 AA; 34534 MW; D01C5D1B CRC32;
```

Query Match 10.4%; Score 375; DB 13; Length 304;
Best Local Similarity 32.1%; Pred. No. 2.57e-54;
Matches 94; Conservative 72; Mismatches 104; Indels 23; Gaps 16;

```
Db 28 NONITEVDAFDRRTFDPAEQYKMNHORRGVALIFNHEHFWHLRLPDRTGLADRNK 87
QY 214 EEDMNFVDAPTISRVEDEKMYRNSSPRGMCLINNEHF--E-QMPTRNGTKADKDLT 270
Db 88 RSLTDLGFVRIFDILKAEDVLKKVFEASRDYSNADCFVCFVLSHGENDHYAYDA-QI 146
QY 271 NLFRCMGYIVICKDNLTRGMLLTIRDFAKHE-SHGDSAILVILSHGEENVIGVDDIPT 329
Db 147 KIETITNMFGRDKCQSLVGKPKIFIIQACRGDKHDD--PVLQ-DSVDS--K--D--ET 196
QY 330 STHEIYDLNANAPRLANKPKIYVQACRGERDNGFPVLDSDVGPAPFLRGMWNRDG 389
Db 197 TV-NQTE-V--DAAGVY-TLPAGADFMICYSAQGYFSHRETVNGSWYIDLCALGKHG 251
QY 390 PLFNLGCVRPQOVWRKKPSQADILIRYATTAQYVSWRNSARGSWFIQAVCEVSTHA 449
Db 252 SLEFTELLTVNKRKVSHRKVDICRDINAIGKKQIPCFASMLTKLYFHPKSK 304
QY 450 KDMVYVELLTVENKKVAC-GFQTSQGSN-I-LKQMPENTSRLLKKFTFWPEAR 499
```

RESULT 3

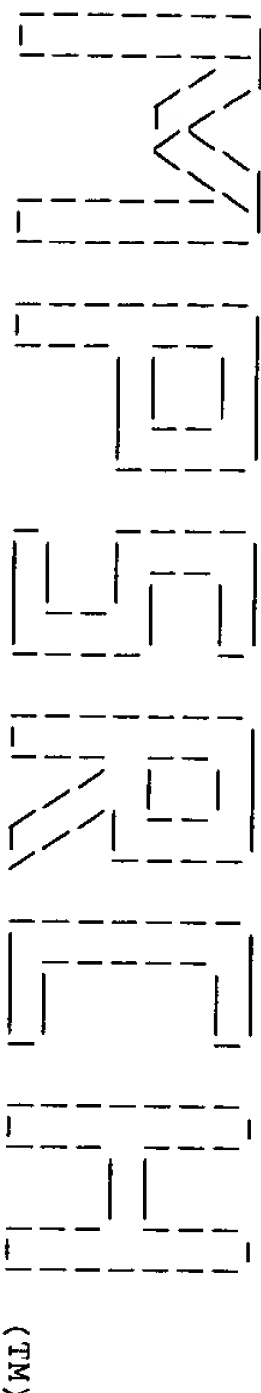
```
ID 018203 PRELIMINARY; PRT; 642 AA.
AC 018203;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE Y48E1B.13 PROTEIN.
GN Y48E1B.13.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEAE; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA MCMURRAY A.;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONEFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL NATURE 368:32-38(1994).
DR EMBL: Z93393; E1354638; -.
SQ SEQUENCE 642 AA; 73590 MW; B3351DE0 CRC32;
```

Query Match 9.9%; Score 356; DB 5; Length 642;
Best Local Similarity 47.2%; Pred. No. 3.92e-50;
Matches 60; Conservative 23; Mismatches 40; Indels 4; Gaps 4;

```
Db 363 YEMNSNPRGTVLLSNENFKMERRVGTGKODEVNTKLFQKLYTVICKRNLEAESMLEA 422
QY 235 YRNFSPPRGMLLINNEHFQMPTRNGTKADKDNLTNLFRCMGYIVICKDNLTRGMLLT 294
Db 423 IKEFA-EMARTDSILFLSHGDGAGSVFCIDDMPVNMEVSTYL-AYH-QNLLKPKWV 479
QY 295 IRDFAKHESHGDSAILVILSHGE-ENVIGVDDIPISTHEIYDLNANAPRLANKPKIV 353
Db 480 AVSACRG 486
QY 354 FVQACRG 360
```

```
RESULT 4
ID 055194 PRELIMINARY; PRT; 452 AA.
AC 055194;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE NEDD2/ICH-1.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98087427.
RA SATO N., MILLIGAN C.E., UCHIYAMA Y., OPPENHEIM R.W.;
RT "Cloning and expression of the cDNA encoding rat caspase-2."
RL GENE 202:127-132(1997).
DR EMBL: U77933; G2769706; -.
DR PROSITE; PS01121; CASPASE_HIS; 1.
DR PROSITE; PS01122; CASPASE_CYS; 1.
SQ SEQUENCE 452 AA; 50728 MW; 0FA96CD3 CRC32;
```

Query Match 9.8%; Score 354; DB 11; Length 452;
Best Local Similarity 33.8%; Pred. No. 1.07e-49;



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 10 11:39:02 1999; Maspar time 31.00 seconds

Tabular output not generated. 885.516 Million cell updates/sec

Title: >US-08-287-669-19
Description: (1-503) from US08287669.pep
Perfect Score: 3603
Sequence: 1 MMRDRSLRLNIMFSSH.....MTSRLKKFYFWPEARNSAV 503

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database: sptrembl9
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 49.838; Variance 85.439; scale 0.583

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	3307	91.8	495	5	002229 C48D1.2 PROTEIN.	0.00e+00
2	375	10.4	304	13	093415 CASPASE-6.	2.57e-54
3	356	9.9	642	5	018203 Y48E1B.13 PROTEIN.	3.92e-50
4	354	9.8	452	11	055194 NEDD2/ICH-1.	1.07e-49
5	350	9.7	277	11	035397 CASPASE 6.	8.07e-49
6	328	9.1	303	11	088550 CASPASE-7.	4.93e-44
7	298	8.3	283	13	042284 CASPASE-1.	1.35e-37
8	295	8.2	283	13	093417 CASPASE-3.	5.87e-37
9	269	7.5	182	6	077623 CASPASE-3 (FRAGMENT).	1.73e-31
10	268	7.4	377	4	075601 EVOLUTIONARILY RELATED	2.80e-31
11	257	7.1	257	11	089094 CASPASE-14.	5.33e-29
12	219	6.1	480	11	089110 CASPASE-8.	2.63e-21
13	207	5.7	422	11	035669 FLICE PARTIAL CDS (FLI	6.03e-19
14	206	5.7	464	4	014676 APOPTOTIC CASPASE MCH5	9.44e-19
15	202	5.6	366	5	002433 DREDD.	5.65e-18
16	202	5.6	488	5	076798 DREDD ISOFORM GAMMA.	5.65e-18
17	202	5.6	494	5	076797 DREDD ISOFORM DELTA.	5.65e-18
18	142	3.9	1238	11	088879 APOPTOTIC PROTEASE ACT	5.18e-07
19	122	3.4	470	5	022919 COSMID C37C3.	9.25e-04
20	114	3.2	445	4	014673 FLAME-1.	1.53e-02

21	114	3.2	449	4	043618 I-FLICE ISOFORM 3.	1.53e-02
22	114	3.2	462	4	060458 USURPIN-BETA.	1.53e-02
23	114	3.2	480	4	015356 CASH ALPHA PROTEIN.	1.53e-02
24	114	3.2	480	4	015519 CASPER.	1.53e-02
25	112	3.1	210	14	066677 ORF E10.	3.03e-02
26	111	3.1	323	11	088602 GAMMA-2 SUBUNIT OF VOL	4.24e-02
27	113	3.1	348	4	043620 I-FLICE ISOFORM 5.	2.15e-02
28	110	3.1	433	2	052401 THIOESTERASE.	5.93e-02
29	112	3.1	442	4	043619 I-FLICE ISOFORM 4.	3.03e-02
30	112	3.1	480	4	015137 CASPASE-LIKE APOPTOSIS	3.03e-02
31	107	3.0	744	10	065375 F12F1.9 PROTEIN.	1.60e-01
32	103	2.9	423	5	P90904 K02B12.7 PROTEIN.	5.84e-01
33	103	2.9	495	2	052406 PUTATIVE 54.5 KDA PROT	5.84e-01
34	106	2.9	703	6	P79122 PININ.	2.22e-01
35	104	2.9	1095	4	060300 KIAA0553 PROTEIN (FRAG	2.22e-01
36	106	2.9	1288	4	015038 KIAA0324 (FRAGMENT).	2.22e-01
37	106	2.9	1791	4	060382 KIAA0324 (FRAGMENT).	2.22e-01
38	101	2.8	328	2	050356 TRANSPOSASE.	1.10e+00
39	101	2.8	330	10	043041 CHALCONE SYNTHASE (EC	1.10e+00
40	102	2.8	358	4	076004 ATP(GTP)-BINDING PROTE	8.02e-01
41	101	2.8	424	5	017469 B0284.2 PROTEIN.	1.10e+00
42	101	2.8	452	10	023146 HNRNP-LIKE PROTEIN.	1.10e+00
43	102	2.8	1645	5	099140 BETA-H SPECTRIN (FRAG	8.02e-01
44	101	2.8	2761	5	018447 CODED FOR BY C. ELEGAN	1.10e+00
45	101	2.8	2810	5	020456 HUM-4 PROTEIN.	1.10e+00

ALIGNMENTS

RESULT ID	1	PRELIMINARY;	PRT;	495 AA.
AC	002229;			
DT	01-JUL-1997 (TREMBLREL. 04, CREATED)			
DT	01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)			
DT	01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)			
DE	C48D1.2 PROTEIN.			
GN	CAENORHABDITIS ELEGANS.			
OS	CAENORHABDITIS ELEGANS.			
OC	EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;			
OC	RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	BURTON J.;			
RL	SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 94150718.			
RA	WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,			
RA	BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,			
RA	CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,			
RA	GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,			
RA	JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,			
RA	LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,			
RA	PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,			
RA	SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,			
RA	THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,			
RA	WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,			
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.			
RT	elegans.";			
RL	NATURE 368:32-38(1994).			
DR	EMBL; Z81049; E1344743; -.			
DR	PROSITE; PS01121; CASPASE_HIS; 1.			
DR	PROSITE; PS01122; CASPASE_CYS; 1.			
SO	SEQUENCE 495 AA; 55549 MW; 7DC9ED9D CRC32;			

Query Match 91.8%; Score 3307; DB 5; Length 495;
Best Local Similarity 98.5%; Pred. No. 0.00e+00;
Matches 465; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db	1	MMRDRSLRLNIMFSSHLKVDLEVLIAKOVLSNDGDMNSCGTVREKRREIVKAV 60
OY	2	MMRDRSLRLNIMFSSHLKVDLEVLIAKOVLSNDGDMNSCGTVREKRREIVKAV 61

RT and CED-3."?
RL ONCOGENE 13:749-755(1996).
RN [2]
RP SEQUENCE OF 30-241 FROM N.A.
RC TISSUE-Ovary;
RX MEDLINE; 96042508.
RA FLAWS J.A., KUGU K., TREBOVICH A.M., DESANTI A., TILLY K.I.,
RT HIRSHFIELD A.N., TILLY J.L.;
RT "interleukin-1 beta-converting enzyme-related proteases (IRPs) and
RT mammalian cell death: dissociation of IRP-induced oligonucleosomal
RT endonuclease activity from morphological apoptosis in granulosa cells
RT of the ovarian follicle.";
RL ENDOCRINOLOGY 136:5042-5053(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 97184204.
RA NI B., WU X., DU Y., SU Y., HAMILTON-BYRD E., ROCKEY P.K.,
RA ROSTECK P. JR., POIRIER G.G., PAUL S.M.;
RT "Cloning and expression of a rat brain interleukin-1beta-converting
RT enzyme (ICE)-related protease (IRP) and its possible role in
RT apoptosis of cultured cerebellar granule neurons.";
RL J. NEUROSCI. 17:1561-1569(1997).
RN [4]
RP SEQUENCE OF 1-264 FROM N.A.
RA YAKOVLEV A.G.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
CC RESPONSIBLE FOR APOPTOSIS EXECUTION. AT THE ONSET OF APOPTOSIS IT
CC PROTEOLYTICALLY CLEAVES POLY(ADP-RIBOSE) POLYMERASE (PARP) AT A
CC 216-ASP-1-GLY-217 BOND. CLEAVES AND ACTIVATES STEROL REGULATORY
CC ELEMENT BINDING PROTEINS (SREBPS) BETWEEN THE BASIC HELIX-LOOP-
CC HELIX LEUCINE ZIPPER DOMAIN AND THE MEMBRANE ATTACHMENT DOMAIN.
CC CLEAVES AND ACTIVATES CASPASE-6, -7 AND -9 (BY SIMILARITY).
CC -1- SUBUNIT: HETERODIMER OF A 17 KD (P17) AND A 12 KD (P12) SUBUNIT
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, LIVER, AND MUSCLE
CC BUT NOT IN KIDNEY OR TESTIS.
CC -1- DEVELOPMENTAL STAGE: HIGHLY EXPRESSED IN NEURON-ENRICHED REGIONS
CC OF THE DEVELOPING BRAIN, BUT DOWN-REGULATED TO LOW LEVELS IN THE
CC ADULT BRAIN.
CC -1- PTM: CLEAVAGE BY GRANZYME B, CASPASE-6, -8 AND -10 GENERATES THE
CC TWO ACTIVE SUBUNITS. ADDITIONAL PROCESSING OF THE PROPEPTIDES IS
CC LIKELY DUE TO THE AUTOCATALYTIC ACTIVITY OF THE ACTIVATED
CC PROTEASE. ACTIVE HETERODIMERS BETWEEN THE SMALL SUBUNIT OF
CC CASPASE-7 PROTEASE AND THE LARGE SUBUNIT OF CPP32 ALSO OCCUR AND
CC VICE VERSA (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CC CASPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U49930; G1518936; -;
DR EMBL; U34685; G1004371; -;
DR EMBL; U84410; G1814027; -;
DR EMBL; U58656; G1381643; -;
DR PROSITE; PS01121; CASPASE_HIS; 1.
DR PROSITE; PS01122; CASPASE_CYS; 1.
DR PFAM; PF00655; ICE_P10; 1.
DR PFAM; PF00656; ICE_P20; 1.
DR HSSP; P42574; IPAU.
KW HYDROLASE; THIOLESTERASE; ZIMOGEN; APOPTOSIS.
FT PROPEP 1 9
FT PROPEP 10 28 BY SIMILARITY.
FT CHAIN 29 175 APOPAIN P17 SUBUNIT.
FT CHAIN 176 277 APOPAIN P12 SUBUNIT.

FT	ACT_SITE	121	121	BY SIMILARITY.
FT	ACT_SITE	163	163	BY SIMILARITY.
FT	CONFLICT	25	29	KSMDS -> OVD (IN REF. 4).
FT	CONFLICT	170	170	C -> S (IN REF. 2).
FT	CONFLICT	178	178	T -> A (IN REF. 2).
FT	CONFLICT	182	182	M -> V (IN REF. 2).
FT	CONFLICT	187	187	I -> K (IN REF. 2).
FT	CONFLICT	190	190	E -> G (IN REF. 3).
FT	CONFLICT	199	199	T -> S (IN REF. 2).
FT	CONFLICT	211	211	D -> G (IN REF. 2).
FT	CONFLICT	236	236	L -> I (IN REF. 4).
FT	CONFLICT	245	245	T -> M (IN REF. 3).
SO	SEQUENCE	277 AA;	31491 MW;	OE106140 CRC32;

Query Match 7.9%; Score 284; DB 1; Length 277;
Best Local Similarity 33.8%; Pred. No. 1.86e-34;
Matches 91; Conservative 63; Mismatches 76; Indels 39; Gaps 20;

Db	37	YKMDYPE-MGLCIINNKFNHKGSTGMSARNGTDVDAANLRETFMALKYEVRRKNDLT-RE 94
Qy	235	YR-NFSSPRGMCLINNEHEFEQ---MPTRNKTKADKDNLTNLFRCMGYTVICKDNLGRC 290
Db	95	EIMELMDSVSKEDHSKRSSFVCVILSHGDEGVIFGTNG-PVDLKLKTSFFRGDYCRSLTG 153
Qy	291	MLTTRDFAKHESHGD-SAIL-VILSHGEENVIIIGVDIPISTHEIYDLNANAPRLAN 348
Db	154	KPKLFIIQACRG-T-E-----LDC--GIET--DSG--TD-D-----D-MAC-----QKI--- 187
Qy	349	KPKIVEVQACRGERRDNGFVLDSDVGPAPFLRRGMDNRDGPFLNFGCVRRPOVQGVWRK 408
Db	188	-PVEADFLYASTAPGYYSWRNSRDGSWFIQSLCAMLKLYAHKLEFMHILTRVNRKVATE 246
Qy	409	KPSQADILLIRYATTAQYVSWRNSARGSWFIQAVCEVFSTHAKMDVVELLTVNKKVACG 468
Db	247	FESFSLDATFHAKKQIPCIYSMLTKELYF 275
Qy	469	FQT-S-QGS-NILKQMPMTSRLKKFYF 494

Search completed: Tue Aug 10 11:38:44 1999
Job time : 59 secs.

[illegible]

```
CC -1- THE SUBUNITS ARE DERIVED FROM THE PRECURSOR SEQUENCE BY A PROBABLE
CC AUTOCATALYTIC MECHANISM AND PROBABLY BY OTHER CASPASES (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CC CASPASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: D89784; D1014718; -.
DR PROSITE; PS01121; CASPASE_HIS; 1.
DR PROSITE; PS01122; CASPASE_CYS; 1.
DR PFAM; PF00655; ICE_p10; 1.
DR PFAM; PF00656; ICE_p20; 1.
DR HSSP; P42574; 1PAU.
KW HYDROLASE; THIOL PROTEASE; ZMOGEN; APOPTOSIS.
FT PROPEP 1 ? BY SIMILARITY.
FT CHAIN ? 186 APOPAIN P17 SUBUNIT.
FT CHAIN 187 282 APOPAIN P12 SUBUNIT.
FT ACT_SITE 131 131 BY SIMILARITY.
FT ACT_SITE 174 174 BY SIMILARITY.
SQ SEQUENCE 282 AA; 32124 MW; E9608701 CRC32;

Query Match 8.4%; Score 302; DB 1; Length 282;
Best Local Similarity 34.9%; Pred. No. 2.90e-38;
Matches 95; Conservative 59; Mismatches 79; Indels 39; Gaps 18;

Db 42 FAHLQNYRTNYPEMGMLIINNKNFHSNMVAVRNGTDVDALRLHETFTGLGYEVMWCNDQ 101
| | : ||||| : | : ||| : | : || : | : |
QY 229 FDEKTYRNFSPPRGMCIIINNEHE--QMPTNRGTAKADKNLTNLFRCMGYTV-ICKDN 285
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 102 -KSSDIIGRLKIS-EEDHSKRSPFYCAILSHGEDSGICVD-VPIHIKNLTDLFRGDR 158
: : : : : | : : : : | : : : : | : : : : | : : : : |
QY 286 LTGRGMLLTRDFAKHESHGD-SALL-YILSHGEENVIT-GVDDIPISTHEIYDLLNAAN 342
: : : : : | : : : : | : : : : | : : : : | : : : : |
Db 159 CKTLVGPKPIFIQACRG-T-E-----LDS--GI-----ET-DS-----CSEPR- 192
| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
QY 343 APRLANPKIVFVQACRGERRDNRPVLDSVDGPAPFLRRGWMDNRDPLENFLGCVRPVQ 402
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | |
Db 193 EEI-ORIPVEADFLYASTVPGYCSWRDKMDSWFIQSCLKMKIKYLSHELIIQLTCVN 251
::: : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 403 QQVWRKKPSQADILIRYATTAQYVSWRNSARGSWEIFQAVCEVFSTHAKDMDVELLTVN 462
: : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 252 HMVALDFETFH-AK--KOIPCVAVSMLTKSFYE 280
: | : | : : : | : | : : | : | : | : | : | : | : | : |
QY 463 KKVACGFQTSGSNILKOMPMTSRLLKKEYF 494
: | : | : : : | : | : : | : | : | : | : | : | : | : |

RESULT 15
ID ICE3_RAT STANDARD; PRT; 277 AA.
AC P55213; P70543; Q62993; P97699;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE APOPAIN PRECURSOR (EC 3.4.22.-) (CYSTEINE PROTEASE CPP32) (YAMA
DE PROTEIN) (CPP-32) (CASPASE-3) (CASP-3) (SREBP CLEAVAGE ACTIVITY 1)
DE (SCA-1) (LICE) (IRP).
GN CASP3 OR CPP32.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIROGNATHI; MURIDAE; MURINAE; RATTUS.
RA [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96358624.
RA JUAN T.S.-C., MCNIECE I.K., JENKINS N.A., GILBERT D.J., COPELAND N.G.,
RA FLETCHER F.A.;
RT "Molecular characterization of mouse and rat CPP32 beta gene encoding
RT a cysteine protease resembling interleukin-1 beta converting enzyme
```


CC ACTIVATING SOME PROTEINS REQUIRED FOR CELL DEATH OR INACTIVATING
CC PROTEINS NECESSARY FOR CELL SURVIVAL (BY SIMILARITY).
CC -1- SUBUNIT: HETERODIMER OF A SMALL AND A LARGE SUBUNIT (BY
CC SIMILARITY).
CC -1- PTM: THE MATURE PROTEASE CAN PROCESS ITS OWN PROPEPTIDE, BUT NOT
CC THAT OF OTHER CASPASES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CC CASPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF025671; G2582497; -.
DR EMBL; U34684; G1004369; -.
DR PROSITE; PS01121; CASPASE_HIS; 1.
DR PROSITE; PS01122; CASPASE_CYS; 1.
DR PFAM; PF00655; ICE_P10; 1.
DR PFAM; PF00656; ICE_P20; 1.
DR HSSP; P29466; IIBC.
KW HYDROLASE; THIOLESTERASE; APOPTOSIS; ZYMOGEN.
FT NON_TER 1
FT CHAIN 1
FT CHAIN <1 266 CASPASE-2 SUBUNIT 1 (BY SIMILARITY).
FT ACT_SITE 267 >312 CASPASE-2 SUBUNIT 2 (BY SIMILARITY).
FT ACT_SITE 210 210 BY SIMILARITY.
FT ACT_SITE 253 253 BY SIMILARITY.
FT NON_TER 312 312
SQ SEQUENCE 312 AA; 35070 MW; 532B7699 CRC32;

Query Match 8.6%; Score 310; DB 1; Length 312;
Best Local Similarity 40.1%; Pred. No. 5.71e-40;
Matches 55; Conservative 27; Mismatches 50; Indels 5; Gaps 3;

Db 125 YRLQSRRLGLALVMSNVHFTGKDLFRSGGDVDTTLVTLFKLLGYNVHLYDQTAQEM 184
QY 235 YRNFSSPRGMCLINNEHF-E-QMPTRNKTKADKDLNLTNFRMGTYVICKDNLTRGM 291
Db 185 QEKLNFAQLPAHRVTDSCIALVSHGVEGGIYVDGKLLQLQEVFRFLFNANCSLQNK 244
QY 292 LTLTRDFAKHESH-GDSALVILSHGEENVIIIGVDIPISTHEIYDLLNAANAPRLANK 349
Db 245 PKMFFIQACRGDETDRG 261
QY 350 PKIVFVQACRGERRDNG 366

RESULT 13
ID ICE7_HUMAN STANDARD; PRT; 303 AA.
AC P55210; Q13364;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CASPASE-7 PRECURSOR (EC 3.4.22.-) (ICE-LIKE APOPTOTIC PROTEASE 3)
DE (ICE-LAP3) (APOPTOTIC PROTEASE MCH-3) (CMH-1).
GN CASP7 OR MCH3.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A. (ALPHA ISOFORM).
RX MEDLINE; 96139498.
RA DUAN H., CHINNAIYAN A.M., HUDSON P.L., WING J.P., HE W.-W.,
RA DIXIT V.M.;
RT "ICE-LAP3, a novel mammalian homologue of the Caenorhabditis elegans
RT cell death protein Ced-3 is activated during Fas- and tumor necrosis
RT factor-induced apoptosis.";
RL J. BIOL. CHEM. 271:1621-1625(1996).
RN [2]
RP SEQUENCE FROM N.A. (ALPHA ISOFORM).

RC TISSUE=SPLEEN;
RX MEDLINE; 96147144.
RA LIPPE J.A., GU Y., SARNECKI C., CARON P.R., SU M.S.-S.;
RT "Identification and characterization of CPP32/Mch2 homolog 1, a novel
RT cysteine protease similar to CPP32.";
RL J. BIOL. CHEM. 271:1825-1828(1996).
RN [3]
RP SEQUENCE FROM N.A. (ALPHA AND BETA ISOFORMS).
RC TISSUE=T-CELL;
RX MEDLINE; 96105019.
RA FERNANDES-ALNEMRI T., TAKAHASHI A., ARMSTRONG R.C., KREBS J.,
RA FRITZ L., TOMASELLI K.J., WANG L., YU Z., CROCE C.M., SALVESON G.,
RA EARNSHAW W.C., LITWACK G., ALNEMRI E.S.;
RT "Mch3, a novel human apoptotic cysteine protease highly related to
RT CPP32.";
RL CANCER RES. 55:6045-6052(1995).
RN [4]
RP SEQUENCE FROM N.A. (ALPHA AND ALPHA' ISOFORMS).
RC TISSUE=FETAL LUNG, AND FETAL SPLEEN;
RX MEDLINE; 97224489.
RA JUAN T.S.-C., MCNIECE I.K., ARGENTO J.M., JENKINS N.A., GILBERT D.J.,
RA COPELAND N.G., FLETCHER F.A.;
RT "Identification and mapping of Casp7, a cysteine protease resembling
RT CPP32 beta, interleukin-1 beta converting enzyme, and CED-3.";
RL GENOMICS 40:86-93(1997).
RN [5]
RP PROCESSING.
RX MEDLINE; 96353838.
RA FERNANDES-ALNEMRI T., ARMSTRONG R.C., KREBS J., SRINIVASULA S.M.,
RA WANG L., BULTRICH F., FRITZ L.C., TRAPANI J.A., TOMASELLI K.J.,
RA LITWACK G., ALNEMRI E.S.;
RT "In vitro activation of CPP32 and Mch3 by Mch4, a novel human
RT apoptotic cysteine protease containing two FADD-like domains.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:7464-7469(1996).
CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
CC RESPONSIBLE FOR APOPTOSIS EXECUTION. CLEAVES AND ACTIVATES STEROL
CC REGULATORY ELEMENT BINDING PROTEINS (SREBPS). PROTEOLYTICALLY
CC CLEAVES POLY(ADP-RIBOSE) POLYMERASE (PARP) AT A 216-ASP-1-GLY-217
CC BOND. OVEREXPRESSION PROMOTES PROGRAMMED CELL DEATH.
CC -1- SUBUNIT: HETERODIMER OF A 20 KD (P20) AND A 11 KD (P11) SUBUNIT.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG, SKELETAL MUSCLE,
CC LIVER, KIDNEY, SPLEEN AND HEART, AND MODERATELY IN TESTIS. NO
CC EXPRESSION IN THE BRAIN.
CC -1- PTM: CLEAVAGES BY GRANTZME B OR CASPASE-10 GENERATE THE TWO ACTIVE
CC SUBUNITS. PROPEPTIDE DOMAINS CAN ALSO BE CLEAVED EFFICIENTLY BY
CC CPP32 PROTEASE. ACTIVE HETERODIMERS BETWEEN THE SMALL SUBUNIT OF
CC CASPASE-7 AND THE LARGE SUBUNIT OF CPP32 PROTEASE ALSO OCCUR AND
CC VICE VERSA.
CC -1- ALTERNATIVE PRODUCTS: THREE ISOFORMS, ALPHA, BETA, AND ALPHA' ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THE SEQUENCE SHOWN HERE IS THAT
CC OF ALPHA ISOFORM. THE BETA ISOFORM IS NOT PROTEOLYTICALLY ACTIVE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CC CASPASE FAMILY.
CC -1- CAUTION: WHAT WE CALL ALPHA' ISOFORM IS KNOWN IN REF.4 AS BETA,
CC BUT AS BETA IS ALREADY DEFINED IN REF.3 WE HAVE CALLED IT ALPHA'.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39613; G1125073; -.
DR EMBL; U40281; G1167555; -.
DR EMBL; U37448; G1117847; -.
DR EMBL; U37449; G1117849; -.
DR EMBL; U67319; G1894913; -.
DR EMBL; U67320; G1894915; -.
DR MIM; 601761; -.
RP PROSITE; PS01121; CASPASE_HIS; 1.


```

DR PROSITE; PS01121; CASPASE_HIS; 1.
DR PROSITE; PS01122; CASPASE_CYS; 1.
DR PFAM; PF00655; ICE_p10; 1.
DR PFAM; PF00656; ICE_p20; 1.
DR HSSP; P42574; 1PAU.
KM HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS.
FT PROPEP 1 23 BY SIMILARITY.
FT CHAIN 24 198 CASPASE-7 SUBUNIT P20 (BY SIMILARITY).
FT PROPEP 199 206 BY SIMILARITY.
FT CHAIN 207 303 CASPASE-7 SUBUNIT P11 (BY SIMILARITY).
FT ACT_SITE 144 144 BY SIMILARITY.
FT ACT_SITE 186 186 BY SIMILARITY.
FT CONFLICT 10 11 EL -> DW (IN REF. 2).
FT CONFLICT 45 45 A -> T (IN REF. 2).
FT CONFLICT 48 49 VR -> RQ (IN REF. 2).
SQ SEQUENCE 303 AA; 34060 MM; 693294C7 CRC32;

Query Match 8.9%; Score 322; DB 1; Length 303;
Best Local Similarity 35.6%; Pred. No. 1.52e-42;
Matches 96; Conservative 63; Mismatches 75; Indels 36; Gaps 19;

Db 59 LYRMDFOK-MGKCIIINNKNFDKATGMVNRGTDKDAGALFKCFQNIGFEVTVHNDSCSA 117
   :|: | :|: ||||| :|: | |||| | :|: | :|: |
Qy 234 MYR-NFSSPRGMCLIIINHEFEQ---MPTRNCTKADKDNLTNLFRCMGTYVICKDNLTR 289
   :|: | :|: ||||| :|: | ||||| :|: | :|: | :|: |

Db 118 KMQDLRLK-ASEEDHSNSACFACVLLSHGEEDLYGKDGVT-PIKDLTAHFRGDCKTLL 175
   | :| | | :|: || :|: ||||| :|: | :|: | :|: |
Qy 290 GMLLTRDFAKHESHGDSA-IL-VILSHGEENVIIIGVDIPISTHEIYDLLNAANAPRLA 347
   :|: | | | | | :|: | || || :|: | :|: | :|: |

Db 176 EKPKLEFIQACRG-T-E-----LD-DGIQA-----DS--GPI-N-D-I--DANP-RN 211
   :||: |:||||| :|: | || || :|: | :|: | :|: |
Qy 348 NKPKIVFVOACRGERRNGEPVLDSVDGPAPFLRGMWNRDGPLFNFLGCVRPQOVWR 407
   :||: |:||||| :|: | || || :|: | :|: | :|: |

Db 212 KIVEADELFAYSTVPGYYSWRNPKGSWFVQALCSILNEHGKDLIMQLTRVNDRVAR 271
   | :|: ||: | :|: | :|: ||||| :|: | :|: | :|: |
Qy 408 KKPSQADILIRYATTAAQYASWRNSARGSWFIQAVCEVFSTHAKMDVVELLEVNNKKVAC 467
   | :|: ||: | :|: | :|: ||||| :|: | :|: | :|: |

Db 272 HFESQSDDPRENEKKQIPCMVSMILTKEYE 301
   | :|: ||: | :|: | :|: ||||| :|: | :|: | :|: |
Qy 468 GFQT-SQGS--NILKOMPEMTSRLKKFYF 494
   | :|: ||: | :|: | :|: ||||| :|: | :|: | :|: |

RESULT 11
ID ICE_DROME STANDARD; PRT; 339 AA.
AC 001382;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CASPASE PRECURSOR (EC 3.4.22.-) (DRICE).
GN ICE.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RX MEDLINE; 97327558.
RA FRASER A.G., EVAN G.I.;
RT "Identification of a Drosophila melanogaster ICE/CED-3-related
   protease, drice."
RI EMBO J. 16:2805-2813(1997).
CC -!- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
CC RESPONSIBLE FOR APOPTOSIS EXECUTION. ACTS DOWNSTREAM OF RPR.
CC CLEAVES BACULOVIRUS P35 AND LAMIN DMO IN VITRO.
CC -!- SUBUNIT: HETERODIMER OF A 21 KD (P21) AND A 12 KD (P12) SUBUNIT.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AT ALL STAGES WHERE APOPTOSIS
CC OCCURS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CC CASPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```

CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; Y12261; E311193; -.
DR      FLYBASE; FBgn0019972; Ice.
DR      PROSITE; PS01121; CASPASE_HIS; 1.
DR      PROSITE; PS01122; CASPASE_CYS; 1.
DR      PFAM; PF00655; ICE_P10; 1.
DR      PFAM; PF00656; ICE_P20; 1.
DR      HSSP; P42574; 1PAU.
KW      HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS.
FT      PROPEP      1      28      BY SIMILARITY.
FT      CHAIN      29      217      CASPASE SUBUNIT P21 (BY SIMILARITY).
FT      PROPEP      218      230      BY SIMILARITY.
FT      CHAIN      231      339      CASPASE SUBUNIT P12 (BY SIMILARITY).
FT      ACT_SITE    169      169      BY SIMILARITY.
FT      ACT_SITE    211      211      BY SIMILARITY.
SQ      SEQUENCE    339 AA;  37393 MW;  0D309C4B CRC32;

```

Query Match	8.7%;	Score 315;	DB 1;	Length 339;
Best Local Similarity	35.7%;	Pred. No. 4.86e-41;		
Matches	84;	Conservative	61;	Mismatches 59;
			Indels 31;	Gaps 21
Db	93	RGMALIFNHEFEVPTLKSRAGTNVDCENLTRYLKQDLDFEVTYKDCRY-KDILRTI-EY	150	
QY	242	RGMCILINNEHE-Q-MPTRNGTKADKDNLTNLFRCMGTYV-ICKDNLTRGMLLTIRDF	298	
Db	151	SASQNHSDSDCLIVAILSHGEMGYIYA-KDQYKLDNIWSEFTANHCPSLAGKPKLFEIQ	209	
QY	299	AKHESHGDSA-ILV-ILSHGEENVIIIGVDIPISTHEIYDILNAAAPRLANKPKRIVFQ	356	
Db	210	ACQGDR-----LDG--GV-T-MQRS-QT-E-T--D--G-DSS-MS--Y-KIPVHADFL	246	
QY	357	ACRGERRDNGFPVLDYDVGPAFLRRGWDNRDGLFNFLLGCVRPQYQVWRKKPSQADIL	416	
Db	247	IAYSTVPGFYVSWRNTTRGTWFMOSLCAELAAANGKRLDILLTLTFVCQRYAVDFES	301	
QY	417	IRYATTAQYVSWRNSARGSWFIQAVCEVFSTHAKMDVVELLTTEVNNKKVACGFQT	471	

```

RESULT 12
ID ICE2 RAT STANDARD; PRT; 312 AA.
AC P55215; O35398;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CASPASE-2 PRECURSOR (EC 3.4.22.-) (CASP-2) (ICH-1 PROTEASE)
DE (FRAGMENT).
DE CASP2 OR ICH1.
GN RATUS NORVEGICUS (RAT).
OS RATUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE OF 1-257 FROM N.A.
RC TISSUE=KIDNEY CORTEX;
RA KAUSHAL G.P., SINGH A.B., SHAH S.V.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 131-312 FROM N.A.
RC TISSUE=OVARY;
RX MEDLINE; 96042508.
RA FLAWS J.A., KUGU K., TROVICH A.M., DESANTI A., TILLY K.I.,
RA HIRSHFIELD A.N., TILLY J.L.;
RT "interleukin-1 beta-converting enzyme-related proteases (IRPs) and
RT mammalian cell death: dissociation of IRP-induced oligonucleosomal
RT endonuclease activity from morphological apoptosis in granulosa cells
RT of the ovarian follicle.";
RL ENDOCRINOLOGY 136:5042-5053(1995).
CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
CC RESPONSIBLE FOR APOPTOSIS EXECUTION. MIGHT FUNCTION BY EITHER

```


modified and this statement is not removed. Usage by and for commercial
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: U56390; G1336027; -
DR EMBL: U60521; G1532151; -
DR MIM: 602234; -
DR PROSITE; PS01121; CASPASE_HIS; 1.
DR PROSITE; PS01122; CASPASE_CYS; 1.
DR PFAM; PF00619; CARD; 1.
DR PFAM; PF00655; ICE_p10; 1.
DR PFAM; PF00656; ICE_p20; 1.
DR HSSP; P42574; 1PAU.
KW HYDROLASE; THIOLESTERASE; ZMOGEN; APOPTOSIS.
FT PROPEP 1 ?
FT CHAIN ? 315 POTENTIAL.
FT PROPEP 316 330 CASPASE-9 SUBUNIT P35.
FT CHAIN 331 416 CASPASE-9 SUBUNIT P10.
FT ACT_SITE 237 237 BY SIMILARITY.
FT ACT_SITE 287 287 BY SIMILARITY.
FT CONFLICT 28 28 V -> A (IN REF. 2).
FT CONFLICT 32 32 R -> S (IN REF. 2).
FT CONFLICT 96 96 G -> A (IN REF. 2).
FT CONFLICT 197 197 L -> P (IN REF. 2).
FT CONFLICT 221 221 R -> Q (IN REF. 2).
SQ SEQUENCE 416 AA; 46322 MW; 91ACEA8E CRC32;

Query Match 10.3%; Score 372; DB 1; Length 416;
Best Local Similarity 32.2%; Pred. No. 1.92e-53;
Matches 77; Conservative 65; Mismatches 79; Indels 18; Gaps 14;

Db 159 PCGHCILNNVNFCESGRLRTGSNIDCEKLRRLSSLHFMVEVKGLIAKMYLALLE 218
QY 241 PRGMCLINNEHF-EQ-MPTNRGTAKDKDNLNLFRCMGYTVICKDNLTGKMLTIRD 297
Db 219 LARQHGALDCVVLISHGCAASHLQEPGAVGTGCPVSVEKIVNIFNGTSCPSLGK 278
QY 298 FAKHESHG-DSAILVILSHG-EE-NV-I--I-GVDDIPISTHEIYDLNMAAPRLANK 349
Db 279 PKLFIQACGGEQKDHGEV-AST-S-PEDESPG-SNPEPDATPFOEGLRTFDQDAI-S 333
QY 350 PKIVFQACRGERDNGFVLDSYDGPAPFLRGWDNRDGPLNLFGLGVRP--QYQVWR 407
Db 334 SLPTPSDIFVSYSTPFGFVSWRDKSGSWYETLDDIFEQWAHSEDLQSLRLRYANAVS 392
QY 408 KKPSQADILIRYATTAQYVSWRNSARGSWFIQAVCEVFSTHAKMDVVELLTVNKVA 466

RESULT 5
ID ICE2_HUMAN STANDARD; PRT; 435 AA.
AC P42575; P42576;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CASPASE-2 PRECURSOR (EC 3.4.22.-) (CASP-2) (ICH-1 PROTEASE) (ICH-1L/1S).
GN CASP2 OR ICH1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
UC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN 111
RP SEQUENCE FROM N.A.
RC TISSUE=FETAL BRAIN;
RX MEDLINE; 94373811.
RA WANG L., MIURA M., BERGERON L., ZHU H., YUAN J.;
RT "Ich-1, an Ice/ced-3-related gene, encodes both positive and negative
regulators of programmed cell death.";
RT CELL 78:739-750(1994).
CC -I- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
RESPONSIBLE FOR APOPTOSIS EXECUTION. MIGHT FUNCTION BY EITHER
ACTIVATING SOME PROTEINS REQUIRED FOR CELL DEATH OR INACTIVATING
PROTEINS NECESSARY FOR CELL SURVIVAL.
CC -I- SUBUNIT: HETERODIMER OF A SMALL AND A LARGE SUBUNIT (BY

SIMILARITY).
CC -I- TISSUE SPECIFICITY: EXPRESSED IN LARGER AMOUNTS IN THE EMBRYONIC
LUNG, LIVER AND KIDNEY THAN IN THE HEART AND BRAIN. IN THE ADULTS
HIGHER LEVEL EXPRESSION IS SEEN IN THE PLACENTA, LUNG, KIDNEY,
PANCREAS THAN IN THE HEART, BRAIN, LIVER AND SKELETAL MUSCLE.
CC -I- PTM: THE MATURE PROTEASE CAN PROCESS ITS OWN PROPEPTIDE, BUT NOT
THAT OF OTHER CASPASES.
CC -I- ALTERNATIVE PRODUCTS: TWO FORMS OF CASPASE-2, A LONG FORM (ICH-1L)
(SHOWN HERE) AND A SHORT FORM (ICH-1S) ARE PRODUCED BY ALTERNATIVE
SPLICING. THEY DIFFER IN THE N- AND C-TERMINI. THE LONG FORM ACTS
AS A POSITIVE REGULATOR OF APOPTOSIS, WHEREAS THE SHORT FORM
FUNCTIONS AS A NEGATIVE REGULATOR OF APOPTOSIS.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CASPASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: U13021; G537292; -
DR EMBL: U13022; G537294; -
DR MIM: 600639; -
DR PROSITE; PS01121; CASPASE_HIS; 1.
DR PROSITE; PS01122; CASPASE_CYS; 1.
DR PFAM; PF00619; CARD; 1.
DR PFAM; PF00655; ICE_p10; 1.
DR PFAM; PF00656; ICE_p20; 1.
DR HSSP; P29466; 1IBC.
KW HYDROLASE; THIOLESTERASE; APOPTOSIS; ZMOGEN; ALTERNATIVE SPLICING.
FT PROPEP 1 ?
FT CHAIN ? 316 POTENTIAL.
FT CHAIN 317 435 CASPASE-2 SUBUNIT 1 (BY SIMILARITY).
FT ACT_SITE 260 260 CASPASE-2 SUBUNIT 2 (BY SIMILARITY).
FT ACT_SITE 303 303 BY SIMILARITY.
FT ACT_SITE 303 303 BY SIMILARITY.
FT MUTAGEN 303 303 C->S: LOSS OF FUNCTION.
FT MUTAGEN 352 352 A->T: LOSS OF FUNCTION.
FT VARSPLIC 1 14 MISSING (IN SHORT FORM).
FT VARSPLIC 306 435 DETRGGVDDQDGNHAGSPGCEESDAGKEKLPKRLPTRS
DMICGYACLKGTAMRNTRKGSWYIEALAOVESERACDMH
VADMLVKVNALIKDEGVAAGTEFHRCKEMSEYCSLGRH
LYLFGHPPT -> GGAIGSLGHLILFTATASLAL (IN
SHORT FORM).
SQ SEQUENCE 435 AA; 48855 MW; 802C3A3C CRC32;

Query Match 10.3%; Score 371; DB 1; Length 435;
Best Local Similarity 34.7%; Pred. No. 3.19e-53;
Matches 94; Conservative 58; Mismatches 95; Indels 24; Gaps 18;

Db 175 YRLQSRPGLALVLSNVHFTGEKELEFRSGGDVDSHTVTLFKLGYDVHVLG-DQ-TAQ 232
QY 235 YRNFSPPRGMLINNEHF-E-QMPTNRGTAKDKDNLNLFRCMGYTV-I-CKDNLGR 289
Db 233 EMOEKLQNFPAQLPAHRTDSCIVALLSHGVEGATYGVGKLLQLOEVFOLFDMANPSLQ 292
QY 290 GMLLTIRDFAKHESH--GDSAILVILSHGEENVIGVDIPISTHEIYDLNMAAPRLA 347
Db 293 NKPKMFIQACRGDETRG--V-DQODG-KN--HAGSPGCEES--D-AG--KEKLPKM-R 340
QY 348 NKPKIVFQACRGERDNGFVLDSYDGPAPFLRGWDNRDGPLNLFGLGVRPQYQVWR 407
Db 341 L-PTRSDMTCGYACLKGTAMRNTRKGSWYIEALAOVESERACDMHVAADMLVKVNALIKD 399
QY 408 KKPSQADILIRYATTAQYVSWRNSARGSWFIQAVCEVFSTHAKMDVVELLTVNKVA-A 466
Db 400 REGYAPGTEFHRCKEMSEYCSLCHLYLFP 430
QY 467 C-GFQTSQGSNLIKQMPMTSRLKRYFWP 496


```

AC P55212;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CASPASE-6 PRECURSOR (EC 3.4.22.-) (APOPTOTIC PROTEASE MCH-2).
GN CASP6 OR MCH2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=T-CELL;
RX MEDLINE; 95316841.
RA FERNANDES-ALNEMRI T., LITWACK G., ALNEMRI E.S.;
RT "Mch2, a new member of the apoptotic Ced-3/Ice cysteine protease gene
RT family.";
RL CANCER RES. 55:2737-2742(1995).
RN [2]
RP PROCESSING.
RC TISSUE=LYMPHOCYTES;
RX MEDLINE; 97059171.
RA SRINIVASULA S.M., FERNANDES-ALNEMRI T., ZANGRILLI J., ROBERTSON N.,
RA ARMSTRONG R.C., WANG L., TRAPANI J.A., TOMASELLI K.J., LITWACK G.,
RA ALNEMRI E.S.;
RT "The Ced-3/interleukin 1beta converting enzyme-like homolog Mch6 and
RT the lamin-cleaving enzyme Mchalpha are substrates for the apoptotic
RT mediator CPP32.";
RL J. BIOL. CHEM. 271:27099-27106(1996).
CC -I- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
CC RESPONSIBLE FOR APOPTOSIS EXECUTION. CLEAVES POLY(ADP-RIBOSE)
CC POLYMERASE IN VITRO, AS WELL AS LAMINS. OVEREXPRESSION PROMOTES
CC PROGRAMMED CELL DEATH.
CC -I- SUBUNIT: HETERODIMER OF A 18 KD (P18) AND A 11 KD (P11) SUBUNIT.
CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -I- PTM: CLEAVAGES BY CPP32, CASPASE-8 OR -10 GENERATE THE TWO ACTIVE
CC SUBUNITS.
CC -I- ALTERNATIVE PRODUCTS: TWO ISOFORMS, ALPHA (SHOWN HERE) AND BETA
CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE BETA ISOFORM DOES NOT
CC SEEM TO HAVE PROTEOLYTIC ACTIVITY.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CC CASPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U20536; G882254; -
CC EMBL; U20537; G882256; -
CC PROSITE; PS01121; CASPASE_HIS; 1.
CC PROSITE; PS01122; CASPASE_CYS; 1.
CC PFAM; PF00655; ICE_P10; 1.
CC PFAM; PF00656; ICE_P20; 1.
CC HSSP; P42574; 1PAU.
KW HYDROLASE; THIOL PROTEASE; APOPTOSIS; ZYMOGEN; ALTERNATIVE SPLICING.
FT PROPEP 1 23
FT CHAIN 24 179 CASPASE-6 SUBUNIT P18.
FT PROPEP 180 193
FT CHAIN 194 293 CASPASE-6 SUBUNIT P11.
FT ACT_SITE 121 121 BY SIMILARITY.
FT ACT_SITE 163 163 BY SIMILARITY.
FT VARSPIC 14 102 MISSING (IN ISOFORM BETA).
SQ SEQUENCE 293 AA; 33409 MW; 91D2D684 CRC32;

Query Match. 10.4%; Score 373; DB 1; Length 293;
Best Local Similarity 33.4%; Pred. No. 1,16e-53;
Matches 98; Conservative 67; Mismatches 104; Indels 24; Gaps 17;

```

```

QY 214 EEDMNFVADPTISRVEDEKTMRYNSSFPRGMCLINNEHF--E-QMPTRNGTAKDKDNL 270
Db 76 RRFSDLGFEVCKFNDLKAELLKIHVEST-VSHADADCFVCVFLSHGEGNHAYADAK- 133
QY 271 NLFRCMGYTVICKDNLTRGMLLTIRDFAKHESHGDSA-IL-VILSHGEENVYIGVDIP 328
Db 134 IEIQTLTGFEKGDCKSHLVGKPKFIIFIOACRGNQHDV--PVI-PLDGV-D--NQT-EKLD 186
QY 329 ISTHEIYDLNLNANAPRLANKPKPIVEYQACRGERDNGFPVLDSYDGVPAFLRGWDNRD 388
Db 187 -T--NITE-V--DAASVY-TLPAGADEFLMCYSVAEGYSSHRETVNGSWYIQDLCENLGKY 239
QY 389 GPLFNFLGCVRPQVOQVWRKKPSQADLLIRYATTAQYVSWRNSARSGSMFIQAVCEVFSTH 448
Db 240 GSSLFETELLTLVNRKVSQRVRDFCKDPSAIGKKQVPCFASMLTKLHFFPKS 292
QY 449 AKDMDVELLTEVNKKVA--CGFQTSQSGSNILKQMPMTSRLKKFFYFWEA 498

RESULT 4
ID ICE9_HUMAN STANDARD; PRT; 416 AA.
AC P55211; Q92852;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CASPASE-9 PRECURSOR (EC 3.4.22.-) (CASP-9) (ICE-LIKE APOPTOTIC
DE PROTEASE 6) (ICE-LAP6) (APOPTOTIC PROTEASE MCH-6) (APOPTOTIC PROTEASE
DE ACTIVATING FACTOR 3) (APAF-3).
GN CASP9 OR MCH6.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96279246.
RA DUAN H., ORTH K., CHINNAIYAN A.M., POIRIER G.G., FROELICH C.J.,
RA HE W.-W., DIXIT V.M.;
RT "ICE-LAP6, a novel member of the ICE/Ced-3 gene family, is activated
RT by the cytotoxic T cell protease granzyme B.";
RL J. BIOL. CHEM. 271:16720-16724(1996).
RN [2]
RP SEQUENCE FROM N.A., AND PROCESSING.
RC TISSUE=T-CELL;
RX MEDLINE; 97059171.
RA SRINIVASULA S.M., FERNANDES-ALNEMRI T., ZANGRILLI J., ROBERTSON N.,
RA ARMSTRONG R.C., WANG L., TRAPANI J.A., TOMASELLI K.J., LITWACK G.,
RA ALNEMRI E.S.;
RT "The Ced-3/interleukin 1beta converting enzyme-like homolog mch6 and
RT the lamin-cleaving enzyme mch2alpha are substrates for the apoptotic
RT mediator CPP32.";
RL J. BIOL. CHEM. 271:27099-27106(1996).
CC -I- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
CC RESPONSIBLE FOR APOPTOSIS EXECUTION. BINDING OF CASPASE-9 TO APAF-
CC 1 LEADS TO ACTIVATION OF THE PROTEASE WHICH THEN CLEAVES AND
CC ACTIVATES CASPASE-3. PROTEOLYTICALLY CLEAVES POLY(ADP-RIBOSE)
CC POLYMERASE (PARP).
CC -I- SUBUNIT: HETERODIMER OF A 35 KD (P35) AND A 10 KD (P10) SUBUNIT.
CC CASPASE-9 AND APAF1 BIND TO EACH OTHER VIA THEIR RESPECTIVE NH2-
CC TERMINAL CED-3 HOMOLOGOUS DOMAINS IN THE PRESENCE OF CYTOCHROME C
CC AND ATP.
CC -I- TISSUE SPECIFICITY: UBIQUITOUS, WITH HIGHEST EXPRESSION IN THE
CC HEART, MODERATE EXPRESSION IN LIVER, SKELETAL MUSCLE, AND
CC PANCREAS. LOW LEVELS IN ALL OTHER TISSUES.
CC -I- PTM: CLEAVAGES AT ASP-315 BY GRANZYME B AND AT ASP-330 BY CPP32
CC GENERATE THE TWO ACTIVE SUBUNITS. CASPASE-8 AND -10 CAN ALSO BE
CC INVOLVED IN THESE PROCESSING EVENTS.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CC CASPASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```


DR	HSSP; P42574; 1CP3.		
KM	HYDROLASE; THIOL PROTEASE; ZMOGEN; APOPTOSIS; PHOSPHORYLATION.		
FT	CHAIN	1	371
FT			
FT	CHAIN	372	503
FT			
FT	DOMAIN	107	205
FT	ACT_SITE	304	304
FT	ACT_SITE	358	358
FT	MUTAGEN	27	27
FT	MUTAGEN	65	65
FT	MUTAGEN	360	360
FT	MUTAGEN	449	449
FT	MUTAGEN	466	466
FT	MUTAGEN	483	483
FT	MUTAGEN	486	486
SO	SEQUENCE	503 AA;	56702 MW; F6874EEE CRC32;

Query Match	100.0%;	Score 3603;	DB 1;	Length 503;
Best Local Similarity	100.0%;	Pred. No. 0.00e+00;		
Matches	503; Conservative	0; Mismatches	0; Indels	0; Gaps

[illegible]

Db	121	G	T	S	P	T	R	V	H	R	D	S	V	S	S	S	F	T	S	Q	D	I	S	R	A	R	S	R	S	R	A	L	H	S	D	R	H	N	S	S	P	P	V	N	A	F	P	S	Q	E	S	180
QY	121	G	T	S	P	T	R	V	H	R	D	S	V <td>S</td> <td>S</td> <td>S</td> <td>F</td> <td>T</td> <td>S</td> <td>Q</td> <td>D</td> <td>I</td> <td>S</td> <td>R</td> <td>A</td> <td>R</td> <td>S</td> <td>R</td> <td>S</td> <td>R</td> <td>A</td> <td>L</td> <td>H</td> <td>S</td> <td>D</td> <td>R</td> <td>H</td> <td>N</td> <td>S</td> <td>S</td> <td>P</td> <td>P</td> <td>V</td> <td>N</td> <td>A</td> <td>F</td> <td>P</td> <td>S</td> <td>Q</td> <td>E</td> <td>S</td> <td>180</td>	S	S	S	F	T	S	Q	D	I	S	R	A	R	S	R	S	R	A	L	H	S	D	R	H	N	S	S	P	P	V	N	A	F	P	S	Q	E	S	180

Db 181 ANSFTGCSLSGYSSRNRSFSKASGPLOYIFHEEDMNFVDAPITSRVFEDEKTMYNRSS 240
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 181 ANSFTGCSLSGYSSRNRSFSKASGPLOYIFHEEDMNFVDAPITSRVFEDEKTMYNRSS 240

Db	241	PRGMC	LI	INNEH	FEOM	PT	RNGT	KADK	DNLN	LF	RCMG	YTVI	CKDN	LTGR	GM	LLTIR	DE	FAK	3000
QY	241	PRGMC	LI	INNEH	FEOM	PT	RNGT	KADK	DNLN	LF	RCMG	YTVI <td>CKDN</td> <td>LTGR</td> <td>GM</td> <td>LLTIR</td> <td>DE</td> <td>FAK</td> <td>3000</td>	CKDN	LTGR	GM	LLTIR	DE	FAK	3000

Db 301 HESHGDSAILVLISHGEEENVIIIGVDDIPISTHEIYDLLNAANAPRLANKPKIVFVQACRG 360
|||||
QY 301 HESHGDSAILVLISHGEEENVIIIGVDDIPISHTHEIYDLLNAANAPRLANKPKIVFVQACRG 360
|||||

[illegible]

Db 421 TTAQYVSWRNSARGSWFIQAVCEVEFSTHAKMDVVELLTEVNKKYACGFQTSQGSNILEK 480
|||||
QY 421 TTAQYVSWRNSARGSWFIQAVCEVEFSTHAKMDVVELLTEVNKKYACGFQTSQGSNILEK 480

```
Db      481  MEEMTSRLKKFFWPEARNsAV 503
        |||
QY      481  MEEMTSRLKKFFWPEARNsAV 503
```

RESULT	2	
ID	CED3_CAEVU	STANDARD; PRT; 496 AA.

DT	01-NOV-1995	(REL. 32, CREATED)
DT	01-NOV-1995	(REL. 32, LAST SEQUENCE UPDATE)
DT	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)
DE	CELL DEATH PROTEIN 3	PRECURSOR (EC 3.4.22.-).

CAENORHABDITIS VULGARIS

CC EUKARYOTA; METAZOA; NECERNITEA; RHABDITIA; RHABDITIDA;
CC RHABDITINA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS

RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE; 94061982.

RA YUAN J., SHAHAM S., LEDOUX S., ELLIS H.M., HORVITZ H.R.;
RT "The C. elegans cell death gene *ced-3* encodes a protein similar to
RT mammalian interleukin-1 beta-converting enzyme."; [PMID: 7564165](#)
RL CELL 75:641-652(1993).

CC -!- FUNCTION: ACTS AS A CYSTEINE PROTEASE IN CONTROLLING PROGRAMMED
CC CELL DEATH BY PROTEOLYTICALLY ACTIVATING OR INACTIVATING A
CC SUBSTRATE PROTEIN OR PROTEINS, A POTENTIAL SUBSTRATE MAY BE CED-4

CC ALTERNATIVELY IT MIGHT DIRECTLY CAUSE CELL DEATH BY
CC PROTEOLYTICALLY CLEAVING PROTEINS THAT ARE CRUCIAL FOR CELL
CC VIABILITY (BY SIMILARITY).

CC -1- SUBUNIT: COULD BE A HETERODIMER OF TWO SUBUNITS DERIVED FROM THE
CC PRECURSOR SEQUENCE BY A PROBABLE AUTOCATALYTIC MECHANISM.
CC -1- PTM: MAY BE REGULATED BY PHOSPHORYLATION.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CC CASPASE FAMILY.
CC
DR PROSITE; PS01121; CASPASE_HIS; 1.

```
DR PROSITE; PS01122; CASPASE_CYS; 1.
DR PFAM; PF00619; CARD; 1.
DR PFAM; PF00655; ICE_p10; 1.
```

DR PFAM; PF00656; ICE_p20; 1.
DR HSP; P42574; 1CP3.
KM HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS; PHOSPHORYLATION

FT	CHAIN	1	364	CELL DEATH PROTEIN 3 SUBUNIT 1 (POTENTIAL).
FT	CHAIN	365	496	CELL DEATH PROTEIN 3 SUBUNIT 2 (POTENTIAL).

FT		(POTENTIAL).
FT	308	308
ACT_SITE		BY SIMILARITY.
FT	351	351
ACT_SITE		BY SIMILARITY.
FT	351	351
ACT_SITE		BY SIMILARITY.

Query Match	83.38;	Score 3001;	DB 1;	Length 496;
Post-Local Similarity	84.39;	Prod No 0	00e+00.	
Seq	SEQUENCE	490 AA;	353943 MW;	CBDFE04D CRC32;

Query Match	83.3%;	Score 3001;	DB 1;	Length 496;
Best Local Similarity	84.3%;	Pred. No. 0.00e+00;		
Matches 425;	Conservative	42;	Mismatches 24;	Indels 13;
				Gaps 8

Db 1 MMQRDRSLERNILVFSSNKLQSEQIILDVLIADKVLNSDNGDMINSCRTERDKRKKEQYKA 60
 ||||| :||: ||:|||||:||||| ||:| ||
 1 MMQRDRSLERNIMFSSHLKVDLEVLIAKQVLNSDNGDMINSCGTVREKREIYKA 60
 QY

```

Db      61  VQRRGDEAFDRFYDALRDTGHELLAAVYEPPLA-T-D-----LQCPMSPASHRRSRALSP 113
        ||||| ||| ||||| ||||| ||||| : ||||| ||||| |||||
QY      61  VQRRGDVAFDAFYDALRSTGHEGLAEVYEPPLARSVDSNAVEFECPMSPASHRRSRALSPA 120

```

D6 114 TFSSPTRVHRDSSVSSTSTYQDVYTPARSTSRSSRPLHTSDRHNVSPS-NSRPSPQ 172
:::|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 GYTSPTRVHRDSSVSSTSTS-YQDIYSRAKRSRS-RALHSSDRHNYSPPVNAPSPQ 178

D6 173 SSANGSFTGCSSTLGYSSSRTRSYSKASASHCIIFHEEDMNVAPTIHRVFDEKMTYRNF 232
|||:|||::|||::|||::|||::|||::|||
QY 179 SSANSSFTGCSSLGYSSSRNRFSKASGPQYIFHEEDMNFPDAPTISR VFDEKMTYRNF 238

```

Db      233  STPRGLCLINNEHEQMPTRNGTKADKNLTNIFRCMGYTVICKDNISGRML-TIRDF 291
      1:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY      239  SSPRGMLINNEHEQMPTRNGTKADKNLTNLFRCMGYTVICKDNLTGRLMLTIRDF 298

```

D0 292 AKNETHGDSAILVILSHGEENVIGVDVSYNVNHEIYDLLNAANPRLANKPKLVFVQAC 351
||::|||||:|||||:|||||:|||||:|||||
QY 299 AKHESHGDSAILVLISHGEENVIGVDIPISTHEIYDLLNAANPRLANKPKIVFVQAC 358

QY 359 RGERDNGFPLDSYDGPAPFLRRGWMDNRDGPLFNFLLGCVRRPQYQYQYWRKKKPSQADILIR 418

Db 410 YATTAQYVSWRNSARGSWFIQAVCEVFSTHAKDMDVELLTEVNKKVACGFGQTSQGANIL 463
|||||:|||||
QY 419 YATTAQYVSWRNSARGSWFIQAVCEVFSTHAKDMDVELLTEVNKKVACGFGQTSQGSNIL 478

DATE	DESCRIPTION	AMOUNT	CHECK NO.	BANK	INITIALS
1/1/20	DEPOSIT	100.00			
1/15/20	PAYROLL	50.00	101		
1/31/20	RENT	25.00	102		
2/1/20	DEPOSIT	150.00			
2/15/20	PAYROLL	50.00	103		
2/28/20	RENT	25.00	104		
3/1/20	DEPOSIT	100.00			
3/15/20	PAYROLL	50.00	105		
3/31/20	RENT	25.00	106		
4/1/20	DEPOSIT	150.00			
4/15/20	PAYROLL	50.00	107		
4/30/20	RENT	25.00	108		
5/1/20	DEPOSIT	100.00			
5/15/20	PAYROLL	50.00	109		
5/31/20	RENT	25.00	110		
6/1/20	DEPOSIT	150.00			
6/15/20	PAYROLL	50.00	111		
6/30/20	RENT	25.00	112		
7/1/20	DEPOSIT	100.00			
7/15/20	PAYROLL	50.00	113		
7/31/20	RENT	25.00	114		
8/1/20	DEPOSIT	150.00			
8/15/20	PAYROLL	50.00	115		
8/31/20	RENT	25.00	116		
9/1/20	DEPOSIT	100.00			
9/15/20	PAYROLL	50.00	117		
9/30/20	RENT	25.00	118		
10/1/20	DEPOSIT	150.00			
10/15/20	PAYROLL	50.00	119		
10/31/20	RENT	25.00	120		
11/1/20	DEPOSIT	100.00			
11/15/20	PAYROLL	50.00	121		
11/30/20	RENT	25.00	122		
12/1/20	DEPOSIT	150.00			
12/15/20	PAYROLL	50.00	123		
12/31/20	RENT	25.00	124		

QY 479 KOMPENTSRLLKKFYFWPEARNSA 502

RESULT	3	
ID	ICE6_HUMAN	STANDARD; PRT; 293 AA

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

(TM)

MPSrch_PP protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Aug 10 11:37:45 1999; MasPar time 15.37 Seconds
925.320 Million cell updates/sec
Tabular output not generated.

Title: >US-08-287-669-19
Description: (1-503) from US08287669.pep
Perfect Score: 3603
Sequence: 1 MMRODRSLERNIMFSSH.....MTSRLKKFYFWPEARNSAV 503

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 51.323; Variance 87.689; scale 0.585

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	3603	100.0	503	1	CED3-CAEEL CELL DEATH PROTEIN 3 P	0.00e+00
2	3001	83.3	496	1	CED3-CAEVU CELL DEATH PROTEIN 3 P	0.00e+00
3	373	10.4	293	1	ICE6-HUMAN CASPASE-6 PRECURSOR (E	1.16e-53
4	372	10.3	416	1	ICE9-HUMAN CASPASE-9 PRECURSOR (E	1.92e-53
5	371	10.3	435	1	ICE2-HUMAN CASPASE-2 PRECURSOR (E	3.19e-53
6	359	10.0	435	1	ICE2-MOUSE CASPASE-2 PRECURSOR (E	1.39e-50
7	358	9.9	276	1	ICE6-MOUSE CASPASE-6 PRECURSOR (E	2.30e-50
8	331	9.2	424	1	ICE2-CHICK CASPASE-2 PRECURSOR (E	1.74e-44
9	328	9.1	303	1	ICE7-MESAU CASPASE-7 PRECURSOR (E	1.52e-42
10	322	8.9	303	1	ICE7-MOUSE CASPASE-7 PRECURSOR (E	4.86e-41
11	315	8.7	339	1	ICE-DROME CASPASE PRECURSOR (E	5.71e-40
12	310	8.6	312	1	ICE2-RAT CASPASE-2 PRECURSOR (E	1.53e-39
13	308	8.5	303	1	ICE7-HUMAN CASPASE-7 PRECURSOR (E	2.90e-38
14	302	8.4	282	1	ICE3-XENLA APOPAIN PRECURSOR (EC	1.86e-34
15	284	7.9	277	1	ICE3-RAT APOPAIN PRECURSOR (EC	3.01e-34
16	283	7.9	277	1	ICE3-HUMAN APOPAIN PRECURSOR (EC	3.01e-34
17	283	7.9	404	1	ICEA-HUMAN INTERLEUKIN-1 BETA CON	7.05e-35
18	283	7.9	521	1	ICEA-HUMAN CASPASE-10 PRECURSOR (E	2.07e-33
19	279	7.7	377	1	ICE4-HUMAN CASPASE-4 PRECURSOR (E	2.07e-33
20	279	7.7	402	1	ILBC-RAT INTERLEUKIN-1 BETA CON	3.36e-33
21	278	7.7	402	1	ILBC-MOUSE INTERLEUKIN-1 BETA CON	3.71e-32
22	273	7.6	299	1	ICE1-SPOFR CASPASE-1 PRECURSOR (E	1.56e-31
23	270	7.5	277	1	ICE3-MOUSE APOPAIN PRECURSOR (EC	

24	272	7.5	418	1	ICE5-HUMAN CASPASE-5 PRECURSOR (E	6.00e-32
25	256	7.1	373	1	ICEB-MOUSE CASPASE-11 PRECURSOR (1.22e-28
26	254	7.0	277	1	ICE3-CRILLO APOPAIN PRECURSOR (EC	3.14e-28
27	234	6.5	419	1	ICEC-MOUSE CASPASE-12 PRECURSOR (3.59e-24
28	231	6.4	386	1	ICEA-XENLA INTERLEUKIN-1 BETA CON	1.43e-23
29	223	6.2	323	1	ICE1-DROME CASPASE-1 PRECURSOR (E	5.62e-22
30	217	6.0	382	1	ICEB-XENLA INTERLEUKIN-1 BETA CON	8.59e-21
31	212	5.9	479	1	ICE8-HUMAN CASPASE-8 PRECURSOR (E	8.20e-20
32	134	3.7	1194	1	APAF-HUMAN APOPTOTIC PROTEASE ACT	1.09e-05
33	120	3.3	272	1	TRA2-DROVI TRANSFORMER-2 SEX-DETE	1.71e-03
34	114	3.2	242	1	PDXJ-ECOLI PYRIDOXAL PHOSPHATE BI	1.34e-02
35	115	3.2	282	1	TRA2-HUMAN TRANSFORMER-2 PROTEIN	9.53e-03
36	115	3.2	376	1	MID2-YEAST MATING PROCESS PROTEIN	9.53e-03
37	111	3.1	272	1	SFR5-HUMAN SPlicing FACTOR, ARGIN	3.63e-02
38	111	3.1	284	1	MT21-STRPN MODIFICATION METHYLASE	3.63e-02
39	112	3.1	506	1	VE2-HPV47 REGULATORY PROTEIN E2.	2.61e-02
40	108	3.0	97	1	YE20-HAEIN HYPOTHETICAL PROTEIN H	9.69e-02
41	108	3.0	221	1	SFR2-CHICK APOPTOSIS REGULATOR CE	1.34e-01
42	107	3.0	280	1	CED9-CAEEL APOPTOSIS REGULATOR CE	1.34e-01
43	107	3.0	604	1	IAP1-HUMAN INHIBITOR OF APOPTOSIS	1.84e-01
44	106	2.9	688	1	STI2-YEAST STE12 PROTEIN.	
45	104	2.9	786	1	STUB-DROME SERINE PROTEINASE STUB	3.47e-01

ALIGNMENTS

RESULT 1
ID CED3-CAEEL STANDARD; PRT; 503 AA.
AC P42573; P45435;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CELL DEATH PROTEIN 3 PRECURSOR (EC 3.4.22.-).
GN CED-3.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94061982.
RA YUAN J., SHAHAM S., LEDOUX S., ELLIS H.M., HORVITZ H.R.;
RT "The C. elegans cell death gene ced-3 encodes a protein similar to
mammalian interleukin-1 beta-converting enzyme.";
RL CELL 75:641-652(1993).
CC -!- FUNCTION: ACTS AS A CYSTEINE PROTEASE IN CONTROLLING PROGRAMMED
CELL DEATH BY PROTEOLYTICALLY ACTIVATING OR INACTIVATING A
SUBSTRATE PROTEIN OR PROTEINS, A POTENTIAL SUBSTRATE MAY BE CED-4.
CC ALTERNATIVELY IT MIGHT DIRECTLY CAUSE CELL DEATH BY
PROTEOLYTICALLY CLEAVING PROTEINS THAT ARE CRUCIAL FOR CELL
VIABILITY.
CC -!- SUBUNIT: COULD BE A HETERODIMER OF TWO SUBUNITS DERIVED FROM THE
PRECURSOR SEQUENCE BY A PROBABLE AUTOCATALYTIC MECHANISM.
CC -!- DEVELOPMENTAL STAGE: MOST ABUNDANT DURING EMBRYOGENESIS AND IS
ALSO DETECTED AT LATER STAGES.
CC -!- PTM: MAY BE REGULATED BY PHOSPHORYLATION.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CASPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L29052; G456417; -.
CC PROSITE; PS01121; CASPASE_HIS; 1.
CC PROSITE; PS01122; CASPASE_CYS; 1.
CC PFAM; PF00619; CARD; 1.
CC PFAM; PF00655; ICE_P10; 1.
CC PFAM; PF00656; ICE_P20; 1.

Page 6

[illegible][illegible]

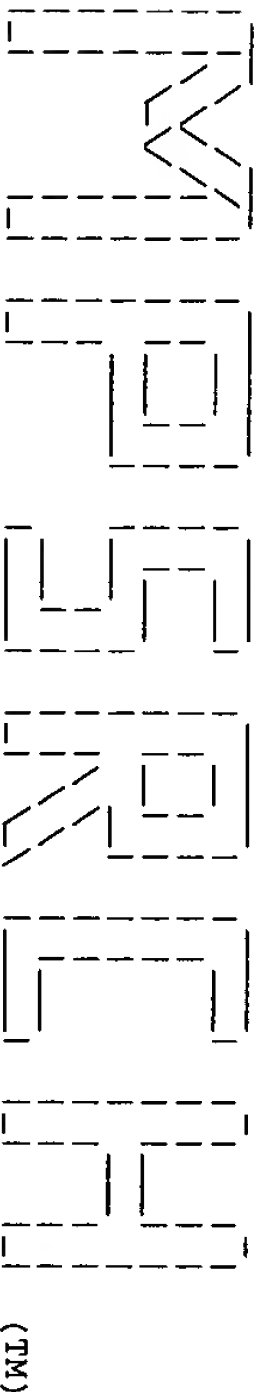
|||||
QY 301 HESHGDSALVILSHGEENVIGVDIPISTHEIYDLNANAPRLANKPKIVFVQACRG 360
SUMMARY
Query Match 10.3%; Score 372; DB 2; Length 416;
Best Local Similarity 32.2%; Pred. No. 2.08e-45;
Matches 77; Conservative 65; Mismatches 79; Indels 18; Gaps 14;
Db 159 PCGHCLINNVCRESGLTRTGSNIDCEKLRFRSSLFHFWYKGDLTAKMVLALLE 218
QY 241 PRGMCLINNENH--EQ-MPTRNGTKADKDNLTNLFRCMGTYVICKDNLTRGMLLTIRD 297
Db 219 LARQDHGALDCVYVILSHGQASHLQFPAVYGTGDCPVSVKIVNIFNGTSCPSLGK 278
QY 298 FAKHESHG-DSAILVILSHG-EE-NV-I---I-GVDDIPISTHEIYDLNANAPRLANK 349
Db 279 PKLFFIQACGGEGKDHGEV-AST-S-PEDESPG-SNPEPDATPFOGLRTFDLDAI-S 333
QY 350 PKIVFVQACRGERDNGFPVLDSDVGPAPFLRRGWDNRDGPLFNLGCVRP--QVQVWR 407
Db 334 SLPTPSDIFVSYTFPGFVSWRDPKSGSWYETLIDIFEQWASEDLQSLLRVANAVS 392
QY 408 KKPQADILIRYATTAQYVSWRNSARGSWFIQAVCEVFSTHAKMDVYELLTEVNKKVA 466
SUMMARY
#molecule_type mRNA
#residues 1-416 #label DUA
#cross-references EMBL:U56390; NID:g1336026; PID:g1336027
#length 416 #molecular-weight 46322 #checksum 797

RESULT 2
ENTRY G02635 #type complete
TITLE ICE-LAP6 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
ACCESSIONS G02635
REFERENCE H01513
#authors Duan, H.; Orth, K.; Chinnaiyan, A.M.; Poirier, G.G.;
#submission submitted to the EMBL Data Library, April 1996
#accession G02635
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-416 #label DUA
#cross-references EMBL:U56390; NID:g1336026; PID:g1336027
#length 416 #molecular-weight 46322 #checksum 797
Query Match 10.3%; Score 372; DB 2; Length 416;
Best Local Similarity 32.2%; Pred. No. 2.08e-45;
Matches 77; Conservative 65; Mismatches 79; Indels 18; Gaps 14;
Db 159 PCGHCLINNVCRESGLTRTGSNIDCEKLRFRSSLFHFWYKGDLTAKMVLALLE 218
QY 241 PRGMCLINNENH--EQ-MPTRNGTKADKDNLTNLFRCMGTYVICKDNLTRGMLLTIRD 297
Db 219 LARQDHGALDCVYVILSHGQASHLQFPAVYGTGDCPVSVKIVNIFNGTSCPSLGK 278
QY 298 FAKHESHG-DSAILVILSHG-EE-NV-I---I-GVDDIPISTHEIYDLNANAPRLANK 349
Db 279 PKLFFIQACGGEGKDHGEV-AST-S-PEDESPG-SNPEPDATPFOGLRTFDLDAI-S 333
QY 350 PKIVFVQACRGERDNGFPVLDSDVGPAPFLRRGWDNRDGPLFNLGCVRP--QVQVWR 407
Db 334 SLPTPSDIFVSYTFPGFVSWRDPKSGSWYETLIDIFEQWASEDLQSLLRVANAVS 392
QY 408 KKPQADILIRYATTAQYVSWRNSARGSWFIQAVCEVFSTHAKMDVYELLTEVNKKVA 466
RESULT 3
ENTRY A54821 #type complete
TITLE Apoptosis regulator ICH-1, stimulatory form L - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 28-Apr-1995 #sequence_revision 28-Apr-1995 #text_change 17-Mar-1999
ACCESSIONS A54821
REFERENCE A54821
#authors Wang, L.; Miura, M.; Bergeron, L.; Zhu, H.; Yuan, J.
#journal Cell (1994) 78:739-750
#title Ich-1, an Ice/ced-3-related gene, encodes both positive and
negative regulators of programmed cell death.
#cross-references MUID:94373811
#accession A54821
#status preliminary
#molecule_type mRNA
#residues 1-435 #label WAN
#cross-references GB:U13021; NID:g537291; PID:g537292

KEYWORDS alternative splicing; apoptosis
SUMMARY
Query Match 10.3%; Score 371; DB 2; Length 435;
Best Local Similarity 34.7%; Pred. No. 3.21e-45;
Matches 94; Conservative 58; Mismatches 95; Indels 24; Gaps 18;
Db 175 YRLQSRPRLALVLSNVHFTGEKELEFRSGGDVDHSTLVTLKLLGYDVHVC-DQ-TAQ 232
QY 235 YRNFSPPRGMLINNENH--E-QMTRNGTKADKDNLTNLFRCMGTYV-I-CKDNLTR 289
Db 233 EMQKLNFAQLPAHRYTDSICVALISHGVEGAIYGVGKLLQLEVFQLFNANCPSLQ 292
QY 290 GMLTTRDFAKHESH--GDSAILVILSHGEENVIGVDIPISTHEIYDLNANAPRLA 347
Db 293 NKPMEFIQACRGDETDRG--V-DQDQD-KN--HAGSPGCEES--D-AG--KEKLPKM-R 340
QY 348 NKPVIVQACRGERDNGFPVLDSDVGPAPFLRRGWDNRDGPLFNLGCVRPQVQVWR 407
Db 341 L-PTRSDMICGYACLKGTAAAMNTRKGSWYIEALQVFSERACDMHVADMLVKYNALIKD 399
QY 408 KKPQADILIRYATTAQYVSWRNSARGSWFIQAVCEVFSTHAKMDVYELLTEVNKKV-A 466
Db 400 REGYAPGTEFHCKEMSEYCTLCRHLYLP 430
QY 467 C-GFQTSQGSNLIKQMPMTSRLKKFYPW 496

RESULT 4
ENTRY JC6507 #type complete
TITLE caspase-2 - Rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 03-Sep-1998 #sequence_revision 03-Sep-1998 #text_change 03-Sep-1998
ACCESSIONS JC6507
REFERENCE JC6507
#authors Sato, N.; Milligan, C.E.; Uchiyama, Y.; Oppenheim, R.W.
#journal Gene (1997) 202:127-132
#title Cloning and expression of the cDNA encoding rat caspase-2.
#accession JC6507
#status preliminary
#residues 1-452 #label SAT
#cross-references GB:U77933
#length 452 #molecular-weight 50728 #checksum 6980
SUMMARY
Query Match 9.8%; Score 354; DB 3; Length 452;
Best Local Similarity 33.8%; Pred. No. 4.96e-42;
Matches 91; Conservative 58; Mismatches 100; Indels 20; Gaps 15;
Db 192 YRLQSORGLALVSNVHFTGEKLEFRSGGDVDHSTLVTLKLLGYDVHVC-DQ-TAQEM 251
QY 235 YRNFSPPRGMLINNENH--E-QMTRNGTKADKDNLTNLFRCMGTYVICKDNLTRGGM 291
Db 252 QEKLNFAQLPAHRYTDSICVALISHGVEGTYGVGKLLQLEVFRLFDNANCPSLQNK 311
QY 292 LITIRDFAKHESH--GDSAILVILSHGEENVIGVDIPISTHEIYDLNANAPRLANK 349
Db 312 PKMFIQACRGDETDRG--V-DQDQD-KNHAQ-S-PGCEES--D-AG--KEKLMKM-RL- 358
QY 350 PKIVFVQACRGERDNGFPVLDSDVGPAPFLRRGWDNRDGPLFNLGCVRPQVQVWRKK 409
Db 359 PTRSDMICGYACLKGNAMNTRKGSWYIEALQVFSERACDMHVADMLVKYNALIKERE 418
QY 410 PSQADILIRYATTAQYVSWRNSARGSWFIQAVCEVFSTHAKMDVYELLTEVNKKV-AC- 467
Db 419 GYAPGTEFHCKEMSEYCTLCQQLYLP 447
QY 468 GFQTSQGSNLIKQMPMTSRLKKFYPW 496

RESULT 5
ENTRY B54821 #type complete



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Aug 10 11:40:55 1999; Maspar time 22.02 Seconds
915.459 Million cell updates/sec
Tabular output not generated.

Title: >US-08-287-669-19
Description: (1-503) from US08287669.pep
Perfect Score: 3603
Sequence: 1 MMRQDRSLRLNIMFSSH.....MTRSLKKEFYFWPEARNSAV 503

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 50.064; Variance 100.638; scale 0.497

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	3591	99.7	503	2	A49429	interleukin-1 beta-co	0.00e+00	
2	372	10.3	416	2	G02635	ICE-LAP6 - human	2.08e-45	
3	371	10.3	435	2	A54821	apoptosis regulator I	3.21e-45	
4	354	9.8	452	3	JC6507	caspase-2 - Rat	4.96e-42	
5	307	8.5	312	2	B54821	apoptosis regulator I	2.41e-33	
6	298	8.3	182	2	I67436	interleukin-1-beta-co	1.05e-31	
7	289	8.0	263	2	C56084	interleukin-1beta con	4.45e-30	
8	285	7.9	277	2	I39005	cysteine proteinase (2.34e-29	
9	283	7.9	277	2	A55315	cysteine proteinase (5.35e-29	
10	286	7.9	311	2	B56084	interleukin-1beta con	1.54e-29	
11	286	7.9	383	2	A56084	interleukin-1beta con	1.54e-29	
12	286	7.9	404	2	A42677	interleukin-1 beta co	1.54e-29	
13	279	7.7	377	2	A57511	interleukin-1 beta co	2.79e-28	
14	278	7.7	399	2	I48911	interleukin-1 beta co	4.22e-28	
15	278	7.7	402	2	A46495	IL-1 beta convertase	4.22e-28	
16	271	7.5	212	2	I67437	cysteine proteinase (7.50e-27	
17	272	7.5	364	2	S62183	cysteine proteinase T	4.97e-27	
18	272	7.5	401	2	A47258	interleukin-1 beta co	4.97e-27	
19	272	7.5	418	2	B57511	interleukin-1 beta co	4.97e-27	
20	267	7.4	277	2	JC5410	CPP32 protein - mouse	3.86e-26	
21	253	7.0	277	2	S64710	cysteine proteinase (1.14e-23	
22	148	4.1	136	2	I53300	interleukin-1-beta-co	1.78e-06	
23	114	3.2	243	2	A42293	pyridoxal phosphate b	8.90e-02	

24	115	3.2	376	2	S52137	MID2 protein - yeast	6.63e-02
25	112	3.1	210	2	S55671	hypothetical protein	1.60e-01
26	111	3.1	272	2	S59042	splicing factor SRP40	2.13e-01
27	111	3.1	284	1	XYSONA	site-specific DNA-met	2.13e-01
28	112	3.1	506	1	W2WL47	E2 protein - human pa	1.60e-01
29	108	3.0	97	2	C64029	hypothetical protein	5.02e-01
30	108	3.0	221	2	B42701	PR264 protein - chick	5.02e-01
31	107	3.0	280	2	A53189	apoptosis suppressor	6.66e-01
32	107	3.0	604	2	S68449	apoptosis inhibitor h	6.66e-01
33	104	2.9	238	2	A57198	splicing factor, argi	1.53e+00
34	103	2.9	471	2	S48265	hypothetical protein	2.02e+00
35	106	2.9	688	2	A33540	transcription factor	8.81e-01
36	104	2.9	786	2	A47547	serine proteinase st	1.53e+00
37	104	2.9	1095	2	T00329	hypothetical protein	1.53e+00
38	106	2.9	1791	2	T02345	hypothetical protein	8.81e-01
39	101	2.8	263	4	I67792	probable glucosylcera	3.46e+00
40	101	2.8	330	2	PS0079	naringenin-chalcone s	3.46e+00
41	101	2.8	424	2	S43560	coiled coil protein B	3.46e+00
42	101	2.8	713	2	S58251	probable membrane pro	3.46e+00
43	102	2.8	923	2	S44664	ZK370.3 protein - Cae	2.65e+00
44	100	2.8	1589	1	RGBC5	cell division control	4.53e+00
45	102	2.8	1645	2	A37792	spectrin beta-H chain	2.65e+00

ALIGNMENTS

RESULT	ENTRY	1	ALIGNMENTS
	TITLE	A49429	#type complete
	ORGANISM	interleukin-1 beta-converting enzyme homolog CED-3 -	
	DATE	07-Apr-1994	#sequence_revision 18-Nov-1994
	ACCESSIONS	A49429	
	REFERENCE	A49429	
	#authors	Yuan, J.; Shaham, S.; Ledoux, S.; Ellis, H.M.; Horvitz, H.R.	
	#journal	Cell (1993) 75:641-652	
	#title	The C. elegans cell death gene ced-3 encodes a protein	
		similar to mammalian interleukin-1 beta-converting enzyme.	
	#cross-references	MUID:94061982	
	#accession	A49429	
	#status	preliminary	
	#molecule_type	DNA	
	##residues	1-503 ##label YUA	
	#note	sequence extracted from NCBI backbone (NCBIN:139825, NCBIF:139826)	
SUMMARY		#length 503 #molecular-weight 56616 #checksum 8929	
	Query Match	99.7%; Score 3591; DB 2; Length 503;	
	Best Local Similarity	99.8%; Pred. No. 0.00e+00;	
	Matches	502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Db	1	MMRQDRSLRLNIMFSSHLKVDLEVLIAKQVLNSDNGDMINSCGTVREKRREIVKA	60
QY	1	MMRQDRSLRLNIMFSSHLKVDLEVLIAKQVLNSDNGDMINSCGTVREKRREIVKA	60
Db	61	VQRRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDSNAVEFECPMSPASHRRSRALSPA	120
QY	61	VQRRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDSNAVEFECPMSPASHRRSRALSPA	120
Db	121	GYSPTVRVHRDSVSSVSTSYQDIYSRARSRSRALHSSDRHNSPPVNAFPSQSS	180
QY	121	GYSPTVRVHRDSVSSVSTSYQDIYSRARSRSRALHSSDRHNSPPVNAFPSQSS	180
Db	181	ANSSFTGCCSLGYSRRNRSFSKASGPTQYIFHEEDMNFVDAPTISRYFDEKTMYNFSS	240
QY	181	ANSSFTGCCSLGYSRRNRSFSKASGPTQYIFHEEDMNFVDAPTISRYFDEKTMYNFSS	240
Db	241	PRGMCIIINNEHEFEQMPTRNGTKADKDLNLNLFRCMGYTVICKDNLTRGMLLTIRDFAK	300
QY	241	PRGMCIIINNEHEFEQMPTRNGTKADKDLNLNLFRCMGYTVICKDNLTRGMLLTIRDFAK	300
Db	301	HESHGDSAILVILSHGEENVITGVDDIPISTHEIYDLNANAPRIANKPKIVFQAACRG	360

PI Horvitz HR, Shahan S, Yuan J;
DR WPI; 94-007551/01.
PT Agents which affect activity of cell death genes - used to
PT develop drugs for treating conditions characterised by cell death
PT or proliferation
PS Claim 99; ; 132pp; English.
CC The sequences given in R45276-328 indicate the positions of possible
CC mutations of the C.elegans cell death gene, ced-3. These mutations
CC occur at conserved amino acid residues of the Ced-3 protein.
CC Fragments of the amino terminal of these proteins act as inhibitors of
CC Ced-3. The ced-3 gene has considerable similarity to human interleukin
CC -1beta convertase (ICE), which converts pro-interleukin-1beta to the
CC active cytokine and is involved in inflammatory response in humans.
CC The similarity between the two sequences suggests that inhibitors of
CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
CC proteins have an overall amino acid similarity of 28%. The ced-3
CC inhibitors may be used for identifying agents which affect the activity
CC of a gene belonging to the to the ced-3/ICE family of genes and for
CC diagnosis of diseases characterised by cell death. They can also be
CC used to develop drugs for treating conditions characterised by cell
CC deaths such as myocardial infarction, stroke, degenerative disease,
CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
CC or drugs for reducing the proliferative capacity or size of a
CC population of cells such as cancerous cells, cells which produce
CC autoreactive antibodies, infected cells, hair follicle cells or cells
CC which are critical to the life of a parasite, pest or recombinant
CC organism. They may also be used in the diagnosis of inflammatory
CC disease.
SQ Sequence 503 AA;

Query Match 99.3%; Score 3579; DB 9; Length 503;
Best Local Similarity 99.4%; Pred. No. 0.00e+00;
Matches 500; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 1 mmrqdrslrlnimfshlkvdeilevliakqvlinsdngdminsctvrekrelvka 60
QY 1 MMRODRSLRLERNIMFSSHLKVDEILEVLI AKOVLNSDNGDMINSCGTVREKREIVKA 60
DB 61 vqrrgdvafdafydairstgheglaevlleplarsvdsnavefecpmshrrsralspa 120
QY 61 VQRRGDVAFDAFYDALIRSTGHEGLAEVLEPLARSVDSNAVEFECPMSPASHRRSRLSPA 120
DB 121 gylsptvrhdsrvssvftsdydiysrarsrsralhssdrhnyssppvna fspqss 180
QY 121 GYLSPTVRHDSRVSSVFTSYDIYSRARSRSRALHSSDRHNYSSPPVNAFPSQSS 180
DB 181 anssftgcslysssrnsrfskasgptylfneedmfvdapltisrvfdektmyrnfs 240
QY 181 ANSSFTGCSLYSSSRNSRFSKASGPTYLFNEEDMNFVDAPLTISRVFDEKTYRNFSS 240
DB 241 prgmcllnehefegmptnrgtkadknltnlfromgytvicknltgrmlltirdfak 300
QY 241 PRGMCLLNEHEFEGMPTNRGTKADKDNLTNLFRCMGYTVICKDNLTGRMLLTIRDFAK 300
DB 301 heshgdsailvlshgeenviigvddipistheiydlinaanprlankpkivfvacrg 360
QY 301 HESHGDSAILVLSHGEENVIIIGVDDIPISTHEIYDLLNAANAPRLANKPKIVFVQACRG 360
DB 361 errdngfpyldsvdgvpaflrrgwandpplfnflgcvrpqvqvwrkpkpsqadiliry 420
QY 361 ERRDNGFPYLDSDVGPAPFLRRGWNDRDPLFNFLGCVRPQVQVWRKPKPSQADILIRYA 420
DB 421 ttaqyvswnsargxwfigavcevfsthakmdvvellevnkvvasafqtsqsnlkq 480
QY 421 TTAQYVSWNSARGSWFIQAVCEVFSTHAKMDVVELLEVNKKVACGFOTSQGSNLLKQ 480
DB 481 mpemtsrlkkfyfwearnsav 503
QY 481 MPEMTSRLKKFYFWPEARNSAV 503

Search completed: Tue Aug 10 11:44:30 1999
Job time : 139 secs.

QY 1 MMRQDRSLLEERNIMFSSHLKVDLEVLIAKQVLNSDNGDMINSCGTVREKREIVKA 60
Db 61 vqrgdvafdafydairstgtgheglaevlleplarsvdsnavefecmpspashrralspa 120
QY 61 VORRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDSNAVEFECMPSPASHRRALSRA 120
Db 121 gytsptrvhrdsvssvssftsytqdiysrarsrsralhsdrhnyssppvnafpsqss 180
QY 121 GYTSPTRVHRDSSVSSVSSFTSYQDIYSRARSRSRALHSSDRHNYSPPVNAFPSQSS 180
Db 181 ansstfgcsslgysssrnsrfskasgptqyifhnedmfvdpatisrvfdektmyrnfs 240
QY 181 ANSSFTGCCSLGYSSSRNRSFSKASGPTQYIFHEDMNFVDAPTISRVEDEKTYRNFSS 240
Db 241 prgmclinnhefegmptnrgtkadkdnltlfrcmgytvickdnltgrgmlltirdfak 300
QY 241 PRGMCLINNHEFEQMPTRNGTKADKDNLTNLFRCMGYTVICKDNLTGRGMLLTIRDFAK 300
Db 301 heshgdaxailvlshgeenviigvddipistheiydlinaanaprlankpkivfvqacrg 360
QY 301 HESHGDSAILVLISHGEENVIIIGVDDIPISTHEIYDLINAAANAPRLANKPKIVFVQACRG 360
Db 361 errdngfpyldsvdgyvpaflrrgwnrdgplfnflgcvrpgvqvwrkksqadillirya 420
QY 361 ERRDNGFPYLDSDVDGVPALFRRGWDNRDGPLFNFLGCVRPQVOQVWRKKPSQADILLIRYA 420
Db 421 ttaqvswrnsaragswfigavcevfsthakmdvvellevnkvasafqtsqgsnllkq 480
QY 421 TTAQYVSWRNSARAGSWFIQAVCEVFSTHAKMDVVELLEVNKVACGFTSQGSNLLKQ 480
Db 481 mpemtsrllkkyfwpearnsav 503
QY 481 MPEMTSRLKKFYFWPEARNSAV 503

RESULT 14
ID R45314 standard; Protein; 503 AA.
AC R45314;
DT 01-JUL-1994 (first entry)
DE Ced-3 mutant A357.
KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
inflammatory response; nematode; diagnosis; myocardial infarction;
stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
pathogenic infection; hair loss; cancer; autoreactive antibody.
KW Caenorhabditis elegans.
OS
FH Key Location/Qualifiers
FT misc_difference 357
FT /label= A357
PN WO9325694-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05705.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-984182.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007551/01.
PT Agents which affect activity of cell death genes - used to
develop drugs for treating conditions characterised by cell death
or proliferation
PS Claim 99; ; 132pp; English.
CC The sequences given in R45276-328 indicate the positions of possible
mutations of the C.elegans cell death gene, ced-3. These mutations
occur at conserved amino acid residues of the Ced-3 protein.
CC Fragments of the amino terminal of these proteins act as inhibitors of
Ced-3. The ced-3 gene has considerable similarity to human interleukin
-1beta convertase (ICE), which converts pro-interleukin-1beta to the
active cytokine and is involved in inflammatory response in humans.
CC The similarity between the two sequences suggests that inhibitors of
ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
proteins have an overall amino acid similarity of 28%. The ced-3
inhibitors may be used for identifying agents which affect the activity
of a gene belonging to the to the ced-3/ICE family of genes and for

CC diagnosis of diseases characterised by cell death. They can also be
CC used to develop drugs for treating conditions characterised by cell
CC deaths such as myocardial infarction, stroke, degenerative disease,
CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
CC or drugs for reducing the proliferative capacity or size of a
CC population of cells such as cancerous cells, cells which produce
CC autoreactive antibodies, infected cells, hair follicle cells or cells
CC which are critical to the life of a parasite, pest or recombinant
CC organism. They may also be used in the diagnosis of inflammatory
CC disease.
SQ Sequence 503 AA;

Query Match 99.3%; Score 3579; DB 9; Length 503;
Best Local Similarity 99.4%; Pred. No. 0.00e+00;
Matches 500; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1 mmrqdrslleernimfsshlkvdlelvliakqvlnsdngdmnscgvtvrekreivka 60
QY 1 MMRQDRSLLEERNIMFSSHLKVDLEVLIAKQVLNSDNGDMINSCGTVREKREIVKA 60
Db 61 vqrgdvafdafydairstgtgheglaevlleplarsvdsnavefecmpspashrralspa 120
QY 61 VORRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDSNAVEFECMPSPASHRRALSRA 120
Db 121 gytsptrvhrdsvssvssftsytqdiysrarsrsralhsdrhnyssppvnafpsqss 180
QY 121 GYTSPTRVHRDSSVSSVSSFTSYQDIYSRARSRSRALHSSDRHNYSPPVNAFPSQSS 180
Db 181 ansstfgcsslgysssrnsrfskasgptqyifhnedmfvdpatisrvfdektmyrnfs 240
QY 181 ANSSFTGCCSLGYSSSRNRSFSKASGPTQYIFHEDMNFVDAPTISRVEDEKTYRNFSS 240
Db 241 prgmclinnhefegmptnrgtkadkdnltlfrcmgytvickdnltgrgmlltirdfak 300
QY 241 PRGMCLINNHEFEQMPTRNGTKADKDNLTNLFRCMGYTVICKDNLTGRGMLLTIRDFAK 300
Db 301 heshgdaxailvlshgeenviigvddipistheiydlinaanaprlankpkivfvqacrg 360
QY 301 HESHGDSAILVLISHGEENVIIIGVDDIPISTHEIYDLINAAANAPRLANKPKIVFVQACRG 360
Db 361 errdngfpyldsvdgyvpaflrrgwnrdgplfnflgcvrpgvqvwrkksqadillirya 420
QY 361 ERRDNGFPYLDSDVDGVPALFRRGWDNRDGPLFNFLGCVRPQVOQVWRKKPSQADILLIRYA 420
Db 421 ttaqvswrnsaragswfigavcevfsthakmdvvellevnkvasafqtsqgsnllkq 480
QY 421 TTAQYVSWRNSARAGSWFIQAVCEVFSTHAKMDVVELLEVNKVACGFTSQGSNLLKQ 480
Db 481 mpemtsrllkkyfwpearnsav 503
QY 481 MPEMTSRLKKFYFWPEARNSAV 503

RESULT 15
ID R45322 standard; Protein; 503 AA.
AC R45322;
DT 01-JUL-1994 (first entry)
DE Ced-3 mutant S435.
KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
inflammatory response; nematode; diagnosis; myocardial infarction;
stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
pathogenic infection; hair loss; cancer; autoreactive antibody.
KW Caenorhabditis elegans.
OS
FH Key Location/Qualifiers
FT misc_difference 435
FT /label= S435
PN WO9325694-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05705.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-984182.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.


```

AC R45262;
DE 01-JUL-1994 (first entry)
KW C.elegans; cell death; gene: ced-3; inhibition; human; parasite;
KW interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
KW inflammatory response; nematode; diagnosis; myocardial infarction;
KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
KW pathogenic infection; hair loss; cancer; autoreactive antibody.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT peptide 1..372
FT /note= "Claim 3, inhibitor fragment"
FT peptide 1..149
FT /note= "Claim 3, inhibitor fragment"
PN WO9325694-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05705.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-984182.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR MPI; 94-007551/01.
DR N-PSDB; Q54401.
PT Agents which affect activity of cell death genes - used to
PT develop drugs for treating conditions characterised by cell death
PT or proliferation
PS Claim 2; Fig 3; 132pp; English.
CC This sequence is encoded by the C.elegans cell death gene, ced-3.
CC Fragments of the amino terminal of this protein act as inhibitors of
CC ced-3. The ced-3 gene has considerable similarity to human interleukin
CC -1beta convertase (ICE), which converts pro-interleukin-1beta to the
CC active cytokine and is involved in inflammatory response in humans.
CC The similarity between the two sequences suggests that inhibitors of
CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
CC proteins have an overall amino acid similarity of 28%. The ced-3
CC inhibitors may be used for identifying agents which affect the activity
CC of a gene belonging to the to the ced-3/ICE family of genes and for
CC diagnosis of diseases characterised by cell death. They can also be
CC used to develop drugs for treating conditions characterised by cell
CC deaths such as myocardial infarction, stroke, degenerative disease,
CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
CC or drugs for reducing the proliferative capacity or size of a
CC population of cells such as cancerous cells, cells which produce
CC autoreactive antibodies, infected cells, hair follicle cells or cells
CC which are critical to the life of a parasite, pest or recombinant
CC organism. They may also be used in the diagnosis of inflammatory
CC disease.
SQ Sequence 503 AA;

```

```

Query Match          99.4%; Score 3583; DB 9; Length 503;
Best Local Similarity 99.6%; Pred. No. 0.00e+00;
Matches 501; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 1 mmrgdrsrllernimmfshlkvdeilevliakqvlinsdngdmnscgtvrekrelvka 60
QY 1 MMRODRKSLERNIMMFSSHLKVDEILEVLIKQVLSNDNGDMINSCGTVREKREIVKA 60
Db 61 vqrrgdvafdaifdalrstgheglavleplarsvdsnavefecpmaspashrralspa 120
QY 61 VORRGDVAFAFDALRSTGHEGLAEVLEPLARSVDSNAVEFECPMASPHRRALSPLA 120
Db 121 gytsptrvhdsyvsstfsydyisrarsrsralhsdrhysppvnafpsqps 180
QY 121 GYTSPTRVHRDSYSSVSTSYQDIYSRARSRSRALHSSDRHNTSPPVNAFPSQPS 180
Db 181 anssftgcslygsssrnsfskasgptqyifhnedmfvdaptisrvfdektmyrnfs 240
QY 181 ANSSFTGCSLYGSSSRNSFSKASGPTQYIFHNEEDMNFVDAPTISRVFDEKTYRNFSS 240
Db 241 prgmclinnhefegmptnrgtkadknltnlfrcmgytvickdnltrgmlltirdfak 300
QY 241 PRGMCLINNHEFEQMPTRNGTKADKDNLTNLFRCMGYTVICKDNLTRGMLLTIRDFAK 300

```

```

Db 301 heshgdsailvlshgeenvligvddipisthetelydlinaanaprlankpivfvqacrg 360
QY 301 HESHGDSAILVLISHGEENVLIGVDDIPISTHETELDLLNAANAPRLANKPIVFVQACRG 360
Db 361 errdngfpvlsvdygypafllrrgwdnrdgplfnflgcvrpqvgvwrkksqadillrya 420
QY 361 ERDNGFPLVDSVDGYPFLRRGWDNRDGPLEFNFLGCVRPQVQVWRKKPSQADILIRYA 420
Db 421 ttaqyvswnrsargswfivagcevfsthakmdvelltvenkkvasafqtsqsnllkq 480
QY 421 TTAQYVSWRNSARGSWFIQAVCEVFSTHAKMDVVELLTVENKKVACGFOQTSQSNILKQ 480
Db 481 mpemtsrlkkfyfwepearnsav 503
QY 481 MPEMTRSRLKKFYFWEPEARNSAV 503

```

```

RESULT 13
ID R45298 standard; Protein; 503 AA.
AC R45298;
DE 01-JUL-1994 (first entry)
KW C.elegans; cell death; gene: ced-3; inhibition; human; parasite;
KW interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
KW inflammatory response; nematode; diagnosis; myocardial infarction;
KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
KW pathogenic infection; hair loss; cancer; autoreactive antibody.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT misc_difference 307
FT /label= S307
PN WO9325694-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05705.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-984182.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR MPI; 94-007551/01.
PT Agents which affect activity of cell death genes - used to
PT develop drugs for treating conditions characterised by cell death
PT or proliferation
PS Claim 99; ; 132pp; English.
CC The sequences given in R45276-328 indicate the positions of possible
CC mutations of the C.elegans cell death gene, ced-3. These mutations
CC occur at conserved amino acid residues of the Ced-3 protein.
CC Fragments of the amino terminal of these proteins act as inhibitors of
CC Ced-3. The ced-3 gene has considerable similarity to human interleukin
CC -1beta convertase (ICE), which converts pro-interleukin-1beta to the
CC active cytokine and is involved in inflammatory response in humans.
CC The similarity between the two sequences suggests that inhibitors of
CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
CC proteins have an overall amino acid similarity of 28%. The ced-3
CC inhibitors may be used for identifying agents which affect the activity
CC of a gene belonging to the to the ced-3/ICE family of genes and for
CC diagnosis of diseases characterised by cell death. They can also be
CC used to develop drugs for treating conditions characterised by cell
CC deaths such as myocardial infarction, stroke, degenerative disease,
CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
CC or drugs for reducing the proliferative capacity or size of a
CC population of cells such as cancerous cells, cells which produce
CC autoreactive antibodies, infected cells, hair follicle cells or cells
CC which are critical to the life of a parasite, pest or recombinant
CC organism. They may also be used in the diagnosis of inflammatory
CC disease.
SQ Sequence 503 AA;

```

```

Query Match          99.3%; Score 3579; DB 9; Length 503;
Best Local Similarity 99.4%; Pred. No. 0.00e+00;
Matches 500; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

Db 1 mmrgdrsrllernimmfshlkvdeilevliakqvlinsdngdmnscgtvrekrelvka 60

```



```
FT misc_difference 283 /note= "possible mutation site"
FT misc_difference 285 /note= "possible mutation site"
FT misc_difference 286 /note= "possible mutation site"
FT misc_difference 287 /note= "possible mutation site"
FT misc_difference 291 /note= "possible mutation site"
FT misc_difference 298 /note= "possible mutation site"
FT misc_difference 304 /note= "possible mutation site"
FT misc_difference 306 /note= "possible mutation site"
FT misc_difference 307 /note= "possible mutation site"
FT misc_difference 310 /note= "possible mutation site"
FT misc_difference 311 /note= "possible mutation site"
FT misc_difference 314 /note= "possible mutation site"
FT misc_difference 315 /note= "possible mutation site"
FT misc_difference 316 /note= "possible mutation site"
FT misc_difference 321 /note= "possible mutation site"
FT misc_difference 323 /note= "possible mutation site"
FT misc_difference 334 /note= "possible mutation site"
FT misc_difference 339 /note= "possible mutation site"
FT misc_difference 344 /note= "possible mutation site"
FT misc_difference 346 /note= "possible mutation site"
FT misc_difference 349 /note= "possible mutation site"
FT misc_difference 350 /note= "possible mutation site"
FT misc_difference 351 /note= "possible mutation site"
FT misc_difference 356 /note= "possible mutation site"
FT misc_difference 357 /note= "possible mutation site"
FT misc_difference 358 /note= "possible mutation site"
FT misc_difference 359 /note= "C385A active site mutation"
FT misc_difference 360 /note= "possible mutation site"
FT misc_difference 371 /note= "possible mutation site"
FT misc_difference 414 /note= "possible mutation site"
FT misc_difference 429 /note= "possible mutation site"
FT misc_difference 434 /note= "possible mutation site"
FT misc_difference 435 /note= "possible mutation site"
FT misc_difference 438 /note= "possible mutation site"
FT misc_difference 449 /note= "possible mutation site"
FT misc_difference 452 /note= "possible mutation site"
FT misc_difference 488 /note= "possible mutation site"
```

```
FT misc_difference 493 /note= "possible mutation site"
FT misc_difference 496 /note= "possible mutation site"
FT misc_difference 496 /note= "possible mutation site"
PN WO9625946-A1.
PD 29-AUG-1996.
PF 23-FEB-1996; U02473.
PR 24-FEB-1995; US-394189.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 96-425082/42.
PT Ced-3 and human interleukin 1-beta convertase genes and proteins
PT useful to treat inflammation and diseases characterised by cell
PT death
PS Claim 60; Fig 3; 139pp; English.
CC Inactivation of the Caenorhabditis elegans Ced-3 cell death protein
CC (R98754), pref. through a substn. of alanine for cysteine at the
CC active site of the protein, provides a mutant Ced-3 protein
CC (R98758) that can prevent programmed cell death in C. elegans
CC when overexpressed using a heat shock promoter. Other mutations
CC at conserved positions can be used to determine amino acid residues
CC important to activity. Corresponding residues in structurally
CC related proteins such as human interleukin-1 beta converting enzyme
CC (R98754) and murine NEDD-2 protein (R98756) may then be altered
CC to produce similar effects.
CC (NB. the active site mutation is stated to be C385A rather than
CC C358A in Claim 62).
SQ Sequence 503 AA;
```

Query Match 99.5%; Score 3586; DB 19; Length 503;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Db 1 mmrgdrslernimmfshlkveilevllakqvlnsdngdminsctgtvrekrrvka 60
QY 1 MMRQDRSLERNIMFSSHLKVEILEVLLAKQVLNSDNGDMINSCGTVREKREIVKA 60
Db 61 vqrgdvafafydalrststghegleveplarsvdsnavfecpmshrrsralspa 120
QY 61 VQRGDVAFDAFYDALRSTGHEGLEVEPLARSVDSNAVEFCPMSPASHRRSRALSPA 120
Db 121 gytsptrvhdsrvssvstsyqdiysrarssrsalhsdsdrhnyssppvnafpsqss 180
QY 121 GYTSPTRVHDSRVSSVSTSYQDIYSRARSSRSRALHSSDRHNYSPPVNAFSPQSS 180
Db 181 ansftgcslysssrnsrfskaspqiyifhedmfvdpatisrvfdektmyrnfs 240
QY 181 ANSFTGCSLYSSSRNRSFSKASGPTQYIFHEEDMNFVDPATISRVFDEKTYRNFS 240
Db 241 prgmclinnhefemprngtkadnltlncrmgytvickdnlgtgrmltirdfak 300
QY 241 PRGMCLINNEHEFEMPTRNKTKADNLTNLCRCMGYTVICKDNLGTGRMLTIRDFAK 300
Db 301 heshgdsailvlshgeenviigvddipistheiydlinaanaprlankpkivfvqaarg 360
QY 301 HESHGDSAILVLISHGEENVIIGVDDIPISTHEIYDLINAANAPRLANKPKIVFVQACRG 360
Db 361 errdngfpvldsvdvpafllrgwdndrdpplnflgcvrpgvqvwrkpsqadiliry 420
QY 361 ERDNGFPVLDSVDVPAFLRRGWDNRDGPPLNFLLGCVRPQVQVWRKPPSQADILIRYA 420
Db 421 ttaqvswrnsargswfigavcefsthakmdvveltevnkvaacgfgtsqsnllkq 480
QY 421 TTAQVSWRNSARGSWFIQAVCEVSTHAKMDVVELTEVNKVAACGFGQTSQGSNLLKQ 480
Db 481 mpemtsrlkkkfyfwearnsav 503
QY 481 MPEMTSRLKKKFYFWEARNSAV 503
```

RESULT 12
ID R45262 standard; Protein; 503 AA.


```

QY 361 ERRDNGFVPLDSVDGVPALRRGWDNRDGPLFNLGCVRPQVQVWRKKPSQADILIRYA 420
DB 421 ttaqyvswnsargswfigavcevfsthakmdvvellevnkkaacgfgtsgsnllkq 480
QY 421 TTAQYVSWRNSARGSWFIQAVCEVFSTHAKMDVVELLEVNKKAACGFGTSGSNILKQ 480
DB 481 mpemtsrllkkfyfwpearnsav 503
QY 481 MPEMTSRLKKFYFWPEARNSAV 503

```

```

RESULT 10
ID R53280 standard; Protein; 503 AA.
AC R53280;
DT 23-JUN-1994 (first entry)
DE Ced-3 (G65R).
KW C. elegans; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT misc_difference 65
FT /label= G65R
FT /note= "From allele n718"
PN WO9325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.
DR N-PSDB; Q64736.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 13; Fig 4; 127pp; English.
CC The sequences given in R53279-88 are encoded by mutant versions of the
CC C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
CC transcript and was most abundant in embryos, but was also detected in
CC larvae and young adults, suggesting that ced-3 is expressed not only
CC in cells undergoing cell death. The four largest introns as well as
CC sequences 5' of the start codon contain repetitive elements, some of
CC which have been characterised in non-coding regions of other C. elegans
CC genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
CC acids in length. Ced-3 is highly hydrophilic with no significant
CC hydrophobic region that might be a transmembrane region. One region
CC of Ced-3 is very rich in serine. It is thought that this region is
CC involved in protein-protein interactions, similar to acid blobs in
CC transcription factors. Of the mutations which occur within the ced-3
CC gene, eight of the mutations are missense mutations, two are nonsense
CC mutations and two are putative splicing mutations. These mutations
CC establish the null phenotype of the ced-3 gene, confirming that ced-3,
CC like ced-4, function is not essential for viability. The ced-3 and
CC ced-4 gene products may be used to develop agents for treating
CC conditions characterised by cell deaths, such as myocardial infarction,
CC stroke, degenerative disease, traumatic brain injury, hypoxia,
CC pathogenic infection, aging or hair loss.
SQ Sequence 503 AA;

```

```

Query Match 99.7%; Score 3591; DB 9; Length 503;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

DB 1 mmrqdrslrllernimmfshlkvdeilevliakgvlnsdngdminsctgtvrekrrreivka 60
QY 1 MMRODRSLRLERNIMMFSSHLKVDLEILEVLIKQVLSNDGDMINSCTGTVERKREIVKA 60
DB 61 vqrrrdvafdafydafrstgheglavleplarsvdsnavefecpmshsralspa 120
QY 61 VQRRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDSNAVEFECPMSPASHRSRALSPA 120

```

```

DB 121 gytsptrvhrdsvsvsftsdyqdiysrararsrslahssdrhnyssppvnafpsqps 180
QY 121 GYTSPTRVHRDSVSVSFTSYQDIYSRARSRSRALHSSDRHNYSSPPVNAFPSQPS 180
DB 181 ansstfgcsslgyssrnrsfskasgptgy1fneedmfvdaptisrvfdektmyrnfs 240
QY 181 ANSSFTGCCSLGYSSRNRSFSKASGPTGY1FHEEDMNFVDAPTISRVFDEKTMYNFSS 240
DB 241 prgmclinnhefegmptnrgtkadkdnltlfrcmgytvickdnltgrmlltirdfak 300
QY 241 PRGMCLINNEHEFEQMPTRNGTKADKDNLTNLFRCMGYTVICKDNLTRGMLLTIRDFAK 300
DB 301 heshgdsailvlshgeenvliqvddipstheiydlinaanaprlankpkivfvqacrg 360
QY 301 HESHGDSAILVILSHGEENVITIGVDDIPSTHEIYDLINAANAPRLANKPKIVFVQACRG 360
DB 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnflgcvrpqvgqwrkksqadiliry 420
QY 361 ERRDNGFVPLDSVDGVPALRRGWDNRDGPLFNLGCVRPQVQVWRKKPSQADILIRYA 420
DB 421 ttaqyvswnsargswfigavcevfsthakmdvvellevnkkaacgfgtsgsnllkq 480
QY 421 TTAQYVSWRNSARGSWFIQAVCEVFSTHAKMDVVELLEVNKKAACGFGTSGSNILKQ 480
DB 481 mpemtsrllkkfyfwpearnsav 503
QY 481 MPEMTSRLKKFYFWPEARNSAV 503

```

```

RESULT 11
ID R98758 standard; Protein; 503 AA.
AC R98758;
DT 17-DEC-1996 (first entry)
DE Nematode Ced-3 cell death protein C358A mutant.
KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;
KW cell death; apoptosis; neural degeneration; inflammation;
KW antiinflammatory.
OS Synthetic.
FH Key Location/Qualifiers
FT region 78..205
FT /label= Serine-rich_region
FT cleavage_site 131
FT /note= "autocleavage site"
FT cleavage_site 371
FT /note= "autocleavage site"
FT misc_difference 183
FT /note= "possible mutation site"
FT misc_difference 234
FT /note= "possible mutation site"
FT modified_site 242
FT /note= "possible mutation site"
FT misc_difference 246
FT /note= "possible mutation site"
FT misc_difference 247
FT /note= "possible mutation site"
FT misc_difference 248
FT /note= "possible mutation site"
FT misc_difference 250
FT /note= "possible mutation site"
FT misc_difference 253
FT /note= "possible mutation site"
FT modified_site 259
FT /note= "possible mutation site"
FT misc_difference 261
FT /note= "possible mutation site"
FT misc_difference 265
FT /note= "possible mutation site"
FT misc_difference 277
FT /note= "possible mutation site"
FT misc_difference 278
FT /note= "possible mutation site"
FT misc_difference 280
FT /note= "possible mutation site"

```


DT 23-JUN-1994 (first entry)
DE Ced-3 (S486F).
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT misc_difference 486
TT /label= S486F
TT /note= "From allele n1163"
PN WO9325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
PI WPI: 94-007542/01.
DR N-PSDB; Q64745.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 13; Fig 4; 127pp; English.
CC The sequences given in R53279-88 are encoded by mutant versions of the
CC C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
CC transcript and was most abundant in embryos, but was also detected in
CC larvae and young adults, suggesting that ced-3 is expressed not only
CC in cells undergoing cell death. The four largest introns as well as
CC sequences 5' of the start codon contain repetitive elements, some of
CC which have been characterised in non-coding regions of other C. elegans
CC genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
CC acids in length. Ced-3 is highly hydrophilic with no significant
CC hydrophobic region that might be a transmembrane region. One region
CC of Ced-3 is very rich in serine. It is thought that this region is
CC involved in protein-protein interactions, similar to acid blobs in
CC transcription factors. Of the mutations which occur within the ced-3
CC gene, eight of the mutations are missense mutations, two are nonsense
CC mutations and two are putative splicing mutations. These mutations
CC establish the null phenotype of the ced-3 gene, confirming that ced-3,
CC like ced-4, function is not essential for viability. The ced-3 and
CC ced-4 gene products may be used to develop agents for treating
CC conditions characterised by cell deaths, such as myocardial infarction,
CC stroke, degenerative disease, traumatic brain injury, hypoxia,
CC pathogenic infection, aging or hair loss.
SQ Sequence 503 AA;
Query Match 99.8%; Score 3595; DB 9; Length 503;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 1 mmrgdrslrlnmmfsshlkvdeilevliakqylnsdngdmnscgtvrekreivka 60
QY 1 MMRODRSLRLERNIMMSSHLKVDEILEVLIKQVLSNDGDMINSCGTVREKREIVKA 60
Db 61 vqrrgdvafdafydairstgtgheglaeleplarsvdsnavefecpmshrrsalspa 120
QY 61 VQRGDVAFDAFYDALRSTGHEGLAEVLEPLARSDNAVEFECPMSPASHRRSRALSPA 120
Db 121 gytsptrvhrdsvsststsydydiysrarsrsralhssdrhnyssppvnaifpsqps 180
QY 121 GYTSPTRVHRDSVSSVSTSTSYDYDIYSRARSRSRALHSSDRHNYSPPVNAIFPSQPS 180
Db 121 gytsptrvhrdsvsststsydydiysrarsrsralhssdrhnyssppvnaifpsqps 180
QY 121 GYTSPTRVHRDSVSSVSTSTSYDYDIYSRARSRSRALHSSDRHNYSPPVNAIFPSQPS 180
Db 181 anssftgcsllysssrnsfskasgptqyifhnedmfvadptisrvfdektmyrnfs 240
QY 181 ANSFTGCSSLGYSSSRNSFSKASGPTQYIFHEEDMNFVADPTISRVFDEKTYRNFS 240
Db 181 anssftgcsllysssrnsfskasgptqyifhnedmfvadptisrvfdektmyrnfs 240
QY 181 ANSFTGCSSLGYSSSRNSFSKASGPTQYIFHEEDMNFVADPTISRVFDEKTYRNFS 240
Db 241 prgmcllinnehfegmptngtkadknltnlfrcmgytvickdnltrgmlltirdfak 300
QY 241 PRGMCLIIINNEHFEGMPTNGTKADKDLTNLFRCMGYTVICKDNLTRGMLLTIRDFAK 300
Db 301 heshgdsailvilshgeenviigvddipistheiydlinaanaprlankpkivfvacrg 360
QY 301 HESHGDSAILVILSHGEENVIIIGVDDIPISTHEIYDLINAANAPRLANKPKIVFVQACRG 360

Db 361 errdngfpvlsvdgvpafllrgwdndrplfnlgcvrpgvqvrwkpkpsqadiliry 420
QY 361 ERRDNGFPLVDSVGPAPAFLLRGWDNRDPLFNLGCVRPQVQVWRKKPSQADILIRYA 420
Db 421 ttaqyvswnsarsgswifgavcevfsthakmdvvelitevnkkaacgfgtsqsnlkg 480
QY 421 TTAQYVSWNSARSGSWFIQAVCEVFSTHAKMDVVELLTVNKKVACGFGFOTSOGSNILKQ 480
Db 481 mpemtfrllkkyfwpearnasav 503
QY 481 MPEMTSRLKKRYFWPEARNASAV 503
RESULT 9
ID R66770 standard; Protein; 503 AA.
AC R66770;
DT 13-SEP-1995 (first entry)
DE Interleukin-1 beta converting enzyme homolog ced3.
KW Interleukin-1 beta converting enzyme; homolog ced3;
KW oncogene bcl-2; programmed cell death; cancer treatment.
OS Caenorhabditis elegans.
PN WO9500160-A.
PD 05-JAN-1995.
PF 10-JUN-1994; U06630.
PR 24-JUN-1993; US-080850.
PA (GENO) GEN HOSPITAL CORP.
PI Miura M, Yuan J;
PI WPI: 95-051742/07.
DR N-PSDB; Q79970.
PT Promoting or preventing programmed cell death in vertebrate cells
PT - by inhibiting the activity of interleukin-1 beta converting
PT enzyme.
PS Example 1; Fig 2C; 116pp; English.
CC Q79970 encodes R66770 interleukin-1 beta converting enzyme homolog
CC ced3, increasing ced3s enzymatic activity can promote the
CC programmed cell death of cancer cells (pref. those overexpressing
CC the bcl-2 oncogene), this can be used as the basis of a new cancer
CC treatment. Alternatively by reducing ced3s enzymatic activity
CC programmed cell death can be inhibited, this may be useful in the
CC development of new cell lines which remain viable in culture for
CC extended or indefinite periods, independent of growth factors.
SQ Sequence 503 AA;
Query Match 99.7%; Score 3593; DB 13; Length 503;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 1 mmrgdrslrlnmmfsshlkvdeilevliakqylnsdngdmnscgtvrekreivka 60
QY 1 MMRODRSLRLERNIMMSSHLKVDEILEVLIKQVLSNDGDMINSCGTVREKREIVKA 60
Db 61 vqrrgdvafdafydairstgtgheglaeleplarsvdsnavefecpmshrrsalspa 120
QY 61 VQRGDVAFDAFYDALRSTGHEGLAEVLEPLARSDNAVEFECPMSPASHRRSRALSPA 120
Db 121 gytsptrvhrdsvsststsydydiysrarsrsralhssdrhnyssppvnaifpsqps 180
QY 121 GYTSPTRVHRDSVSSVSTSTSYDYDIYSRARSRSRALHSSDRHNYSPPVNAIFPSQPS 180
Db 121 gytsptrvhrdsvsststsydydiysrarsrsralhssdrhnyssppvnaifpsqps 180
QY 121 GYTSPTRVHRDSVSSVSTSTSYDYDIYSRARSRSRALHSSDRHNYSPPVNAIFPSQPS 180
Db 181 anssftgcsllysssrnsfskasgptqyifhnedmfvadptisrvfdektmyrnfs 240
QY 181 ANSFTGCSSLGYSSSRNSFSKASGPTQYIFHEEDMNFVADPTISRVFDEKTYRNFS 240
Db 181 anssftgcsllysssrnsfskasgptqyifhnedmfvadptisrvfdektmyrnfs 240
QY 181 ANSFTGCSSLGYSSSRNSFSKASGPTQYIFHEEDMNFVADPTISRVFDEKTYRNFS 240
Db 241 prgmcllinnehfegmptngtkadknltnlfrcmgytvickdnltrgmlltirdfak 300
QY 241 PRGMCLIIINNEHFEGMPTNGTKADKDLTNLFRCMGYTVICKDNLTRGMLLTIRDFAK 300
Db 301 heshgdsailvilshgeenviigvddipistheiydlinaanaprlankpkivfvacrg 360
QY 301 HESHGDSAILVILSHGEENVIIIGVDDIPISTHEIYDLINAANAPRLANKPKIVFVQACRG 360
Db 361 errdngfpvlsvdgvpafllrgwdndrplfnlgcvrpgvqvrwkpkpsqadiliry 420

CC involved in protein-protein interactions, similar to acid blobs in
CC transcription factors. Of the mutations which occur within the ced-3
CC gene, eight of the mutations are missense mutations, two are nonsense
CC mutations and two are putative splicing mutations. These mutations
CC establish the null phenotype of the ced-3 gene, confirming that ced-3,
CC like ced-4, function is not essential for viability. The ced-3 and
CC ced-4 gene products may be used to develop agents for treating
CC conditions characterised by cell deaths, such as myocardial infarction,
CC stroke, degenerative disease, traumatic brain injury, hypoxia,
CC pathogenic infection, aging or hair loss.
SQ Sequence 503 AA;

Query Match 99.8%; Score 3596; DB 9; Length 503;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 502; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 mmrgdrslrlernimfshlkvdeifevliakqvlinsdngdmnscgtvrekreivka 60
|||
QY 1 MMRQDRSLRLERNIMFSSHLKVDLEILVLIKQVLSNDGDMNSCGTVREKREIVKA 60
Db 61 vqrrgdvafdaalrstgheglaeleplarsvdsnavefecpmshrrsralspa 120
|||
QY 61 VQRRGDVAFDAFALRSTGHEGLAEVLEPLARSVDSNAVEFECPMSPASHRRSRLSPA 120
Db 121 gytsptrvhrdsvssvssftsdydysrararsralhssdrhysppvnafpsqps 180
|||
QY 121 GYTSPTRVHRDSVSSVSSFTSYDYDYSRARARSRLHSSDRHNYSPPVNAFPSQPS 180
Db 181 ansftgcslgysssrnsrfskasgptqyifneedmfvdaptisrvfdektmyrnfs 240
|||
QY 181 ANSFTGCSSLGYSSSRNRSFSKASGPTQYIFHEEDMNFVDAPTISRVFDEKTYRNFS 240
Db 241 prgmclinnhefemprtnrgtkadknltnlfrcmgytvickdnltrgmlltirdfak 300
|||
QY 241 PRGMCLINNHEFEMPTRNKTKADKDLTNLFRCMGYTVICKDNLTRGMLLTIRDFAK 300
Db 301 heshgdsailvlshgeenvliigvddipistheiydlinaanaprlankpkivfvqacrg 360
|||
QY 301 HESHGDSAILVLISHGEENVLIIGVDDIPISTHEIYDLINAANAPRLANKPKIVFVQACRG 360
Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnlfgcvrpqvgvwrkkpsqadillrya 420
|||
QY 361 ERRDNGFPVLDSVDGPAPFLRRGWDNRDGLFNFLGCVRPQVQVWRKKPSQADILIRYA 420
Db 421 ttaqyvswnrsaragswfigavcevfsthakmdvvellevnkkyacgfgtsgsnilkq 480
|||
QY 421 TTAQYVSWRNSARCSWFIQAVCEVFSTHAKMDVVELLEVNKKYACGFGTSGSNILKQ 480
Db 481 mpemtsrllkkfyfwepearnsav 503
|||
QY 481 MPEMTSRLKKFYFWPEARNSAV 503

RESULT 7
ID R53287 standard; Protein; 503 AA.
AC R53287;
DT 23-JUN-1994 (first entry)
DE Ced-3 (E483K).
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT misc_difference 483
FT /label= E483K
FT /note= "From allele n2426"
PD W09325685-A.
PN 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.

PI Horvitz HR, Shahan S, Yuan J;
DR WPI; 94-007542/01.
DR N-PSDB; Q64744.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 13; Fig 4; 127pp; English.
CC The sequences given in R53279-88 are encoded by mutant versions of the
CC C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
CC transcript and was most abundant in embryos, but was also detected in
CC larvae and young adults, suggesting that ced-3 is expressed not only
CC in cells undergoing cell death. The four largest introns as well as
CC sequences 5' of the start codon contain repetitive elements, some of
CC which have been characterised in non-coding regions of other C. elegans
CC genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
CC acids in length. Ced-3 is highly hydrophilic with no significant
CC hydrophobic region that might be a transmembrane region. One region
CC of Ced-3 is very rich in serine. It is thought that this region is
CC involved in protein-protein interactions, similar to acid blobs in
CC transcription factors. Of the mutations which occur within the ced-3
CC gene, eight of the mutations are missense mutations, two are nonsense
CC mutations and two are putative splicing mutations. These mutations
CC establish the null phenotype of the ced-3 gene, confirming that ced-3,
CC like ced-4, function is not essential for viability. The ced-3 and
CC ced-4 gene products may be used to develop agents for treating
CC conditions characterised by cell deaths, such as myocardial infarction,
CC stroke, degenerative disease, traumatic brain injury, hypoxia,
CC pathogenic infection, aging or hair loss.
SQ Sequence 503 AA;

Query Match 99.8%; Score 3596; DB 9; Length 503;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 mmrgdrslrlernimfshlkvdeifevliakqvlinsdngdmnscgtvrekreivka 60
|||
QY 1 MMRQDRSLRLERNIMFSSHLKVDLEILVLIKQVLSNDGDMNSCGTVREKREIVKA 60
Db 61 vqrrgdvafdaalrstgheglaeleplarsvdsnavefecpmshrrsralspa 120
|||
QY 61 VQRRGDVAFDAFALRSTGHEGLAEVLEPLARSVDSNAVEFECPMSPASHRRSRLSPA 120
Db 121 gytsptrvhrdsvssvssftsdydysrararsralhssdrhysppvnafpsqps 180
|||
QY 121 GYTSPTRVHRDSVSSVSSFTSYDYDYSRARARSRLHSSDRHNYSPPVNAFPSQPS 180
Db 181 ansftgcslgysssrnsrfskasgptqyifneedmfvdaptisrvfdektmyrnfs 240
|||
QY 181 ANSFTGCSSLGYSSSRNRSFSKASGPTQYIFHEEDMNFVDAPTISRVFDEKTYRNFS 240
Db 241 prgmclinnhefemprtnrgtkadknltnlfrcmgytvickdnltrgmlltirdfak 300
|||
QY 241 PRGMCLINNHEFEMPTRNKTKADKDLTNLFRCMGYTVICKDNLTRGMLLTIRDFAK 300
Db 301 heshgdsailvlshgeenvliigvddipistheiydlinaanaprlankpkivfvqacrg 360
|||
QY 301 HESHGDSAILVLISHGEENVLIIGVDDIPISTHEIYDLINAANAPRLANKPKIVFVQACRG 360
Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnlfgcvrpqvgvwrkkpsqadillrya 420
|||
QY 361 ERRDNGFPVLDSVDGPAPFLRRGWDNRDGLFNFLGCVRPQVQVWRKKPSQADILIRYA 420
Db 421 ttaqyvswnrsaragswfigavcevfsthakmdvvellevnkkyacgfgtsgsnilkq 480
|||
QY 421 TTAQYVSWRNSARCSWFIQAVCEVFSTHAKMDVVELLEVNKKYACGFGTSGSNILKQ 480
Db 481 mpemtsrllkkfyfwepearnsav 503
|||
QY 481 MPEMTSRLKKFYFWPEARNSAV 503

RESULT 8
ID R53288 standard; Protein; 503 AA.
AC R53288;


```

QY      241 PRGMCLINNEHEFQMPTRNGTKADKDNLTNLFRCMGYTVICKDNLTGRGMLLTIRDEAK 300
Db      301 heshgdsailvishgeenviigvddipistheiydlinaanaprlankpkivfgacrg 360
QY      301 HESHGDSAILVISHGEENVIIIGVDDIPISTHEIYDLINAAANAPRLANKPKIVFVQACRG 360
Db      361 errdngfpvldsvdgvpaflrrgwdndrpglfnflgcvrpgvqvwrrkpsqadiliry 420
QY      361 ERRDNGFPVLDSVDGVPAPFLRRGWDNDRGPFLFNFLGCVRPQVQVWRKPKPSQADILIRYA 420
Db      421 ttaqvswrnsarsgswfiqavcevfsthakmdvvellevnkvvvcgfgtsgsnilkq 480
QY      421 TTAQVSWRNSARSGSWFIQAVCEVFSTHAKMDVVELLLEVNKVVCACGFGTSGSNILKQ 480
Db      481 mpemtsrllkkfyfwepearnsav 503
QY      481 MPEMTSRLKKFYFWPEARNSAV 503

RESULT 5
ID      R53281 standard; Protein; 503 AA.
AC      R53281;
DT      23-JUN-1994 (first entry)
DE      Ced-3 (G360S).
KW      C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW      embryogenesis; cell death; hydrophilic; transmembrane; region;
KW      hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW      protein synthesis; null phenotype; calcium-binding domain.
OS      Caenorhabditis elegans.
FH      Key
FT      misc_difference 360
FT      /label= G360S
FT      /note= "From allele n2433"
PN      W09325685-A.
PD      23-DEC-1993.
PF      14-JUN-1993; U05701.
PR      12-JUN-1992; US-897788.
PR      20-NOV-1992; US-979638.
PA      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI      Horvitz HR, Shaham S, Yuan J;
DR      MPI; 94-007542/01.
DR      N-PSDB; Q64737.
DT      Isolated C elegans cell death genes ced-3 and ced-4 - used to
DT      develop agents to increase or prevent cell death in organisms
PS      Claim 13; Fig 4; 127pp; English.
CC      The sequences given in R53279-88 are encoded by mutant versions of the
CC      C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
CC      transcript and was most abundant in embryos, but was also detected in
CC      larvae and young adults, suggesting that ced-3 is expressed not only
CC      in cells undergoing cell death. The four largest introns as well as
CC      sequences 5' of the start codon contain repetitive elements, some of
CC      which have been characterised in non-coding regions of other C. elegans
CC      genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
CC      acids in length. Ced-3 is highly hydrophilic with no significant
CC      hydrophobic region that might be a transmembrane region. One region
CC      of Ced-3 is very rich in serine. It is thought that this region is
CC      involved in protein-protein interactions, similar to acid blobs in
CC      transcription factors. Of the mutations which occur within the ced-3
CC      gene, eight of the mutations are missense mutations, two are nonsense
CC      mutations and two are putative splicing mutations. These mutations
CC      establish the null phenotype of the ced-3 gene, confirming that ced-3,
CC      like ced-4, function is not essential for viability. The ced-3 and
CC      ced-4 gene products may be used to develop agents for treating
CC      conditions characterised by cell deaths, such as myocardial infarction,
CC      stroke, degenerative disease, traumatic brain injury, hypoxia,
CC      pathogenic infection, aging or hair loss.
SQ      Sequence 503 AA;

Query Match      99.8%; Score 3597; DB 9; Length 503;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 502; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db      1 mmrgdrsl1ernimfshlkdvdellevlakgvlnsdngmnsogtvtrekreivka 60

```

```

QY      1 MMRODRRSLLEARNIMMFSSHLKVDLEILEVLIKQVLNSDNGMDMINSCTVREKRREIVKA 60
Db      61 vqrrgdvafafdydalirstgheglaeyleplarsvdsnavefecmpspashrrsralspa 120
QY      61 VQRRGDVAFADFADALIRSTGHEGLAEYLEPLARSVDSNAVEFECMPSPASHRRSRALSPA 120
Db      121 gytsptrvhrdsvssvssftsyqdllyrsarsrsrsralhsdtrhnyssppvnafpsqps 180
QY      121 GYTSPTVRHRSVSSVSSFTSYQDIYSRARSRSRSLALHSSDRHNYSSPPVNAFPSQPS 180
Db      181 anssftgcsslgysssrnsrfskasgptqylfneedmfvvadptisrvfdektmyrnfs 240
QY      181 ANSSFTGCSSLGYSSSRNRSFSKASGPTQYIFNEEDMNFVDAPTISRVEDEKTMRYNFS 240
Db      241 prgmclinnehefemqptngtkadkdnltnlfrcmgytvickdnltgrgmlltirdfak 300
QY      241 PRGMCLINNEHEFQMPTRNGTKADKDNLTNLFRCMGYTVICKDNLTGRGMLLTIRDEAK 300
Db      301 heshgdsailvishgeenviigvddipistheiydlinaanaprlankpkivfgacrg 360
QY      301 HESHGDSAILVISHGEENVIIIGVDDIPISTHEIYDLINAAANAPRLANKPKIVFVQACRG 360
Db      361 errdngfpvldsvdgvpaflrrgwdndrpglfnflgcvrpgvqvwrrkpsqadiliry 420
QY      361 ERRDNGFPVLDSVDGVPAPFLRRGWDNDRGPFLFNFLGCVRPQVQVWRKPKPSQADILIRYA 420
Db      421 ttaqvswrnsarsgswfiqavcevfsthakmdvvellevnkvvvcgfgtsgsnilkq 480
QY      421 TTAQVSWRNSARSGSWFIQAVCEVFSTHAKMDVVELLLEVNKVVCACGFGTSGSNILKQ 480
Db      481 mpemtsrllkkfyfwepearnsav 503
QY      481 MPEMTSRLKKFYFWPEARNSAV 503

RESULT 6
ID      R53279 standard; Protein; 503 AA.
AC      R53279;
DT      23-JUN-1994 (first entry)
DE      Ced-3 (L27F).
KW      C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW      embryogenesis; cell death; hydrophilic; transmembrane; region;
KW      hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW      protein synthesis; null phenotype; calcium-binding domain.
OS      Caenorhabditis elegans.
FH      Key
FT      misc_difference 27
FT      /label= L27F
FT      /note= "From allele n1040"
PN      W09325685-A.
PD      23-DEC-1993.
PF      14-JUN-1993; U05701.
PR      12-JUN-1992; US-897788.
PR      20-NOV-1992; US-979638.
PA      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI      Horvitz HR, Shaham S, Yuan J;
DR      MPI; 94-007542/01.
DR      N-PSDB; Q64735.
DT      Isolated C elegans cell death genes ced-3 and ced-4 - used to
DT      develop agents to increase or prevent cell death in organisms
PS      Claim 13; Fig 4; 127pp; English.
CC      The sequences given in R53279-88 are encoded by mutant versions of the
CC      C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
CC      transcript and was most abundant in embryos, but was also detected in
CC      larvae and young adults, suggesting that ced-3 is expressed not only
CC      in cells undergoing cell death. The four largest introns as well as
CC      sequences 5' of the start codon contain repetitive elements, some of
CC      which have been characterised in non-coding regions of other C. elegans
CC      genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
CC      acids in length. Ced-3 is highly hydrophilic with no significant
CC      hydrophobic region that might be a transmembrane region. One region
CC      of Ced-3 is very rich in serine. It is thought that this region is

```


23-DEC-1993.
 PD 14-JUN-1993; U05701.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-979638.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shahan S, Yuan J;
 DR WPI: 94-007542/01.
 DR N-PSDB; Q64742.
 PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 PT develop agents to increase or prevent cell death in organisms
 PS Claim 13; Fig 4; 127pp; English.
 CC The sequences given in R53279-88 are encoded by mutant versions of the
 CC C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
 CC transcript and was most abundant in embryos, but was also detected in
 CC larvae and young adults, suggesting that ced-3 is expressed not only
 CC in cells undergoing cell death. The four largest introns as well as
 CC sequences 5' of the start codon contain repetitive elements, some of
 CC which have been characterised in non-coding regions of other C. elegans
 CC genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
 CC acids in length. Ced-3 is highly hydrophilic with no significant
 CC hydrophobic region that might be a transmembrane region. One region
 CC of Ced-3 is very rich in serine. It is thought that this region is
 CC involved in protein-protein interactions, similar to acid blobs in
 CC transcription factors. Of the mutations which occur within the ced-3
 CC gene, eight of the mutations are missense mutations, two are nonsense
 CC mutations and two are putative splicing mutations. These mutations
 CC establish the null phenotype of the ced-3 gene, confirming that ced-3,
 CC like ced-4, function is not essential for viability. The ced-3 and
 CC ced-4 gene products may be used to develop agents for treating
 CC conditions characterised by cell deaths, such as myocardial infarction,
 CC stroke, degenerative disease, traumatic brain injury, hypoxia,
 CC pathogenic infection, aging or hair loss.
 SQ Sequence 503 AA;

Query Match 99.9%; Score 3599; DB 9; Length 503;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 mmrgdrslrllernimmfshlkvdeilevliakqvlinsdngdmnscgtvrekrrivka 60
 QY 1 MMRODRSLRLERNIMMFSSHLKVEILEVLIKQVLNSDNGDMINSCGTVERKREIVKA 60
 Db 61 vqrrgdvafadafdalrstqheglavleplarsvsnavefecmpspashrralspa 120
 QY 61 VORRGDVAFAFDALRSTGHEGLAEVLEPLARSVSNAVEFECMPSPASHRRALSPA 120
 Db 121 gytsptrvhrdsrvssvssftsdyqdiysrarsrsralhssdrhnysspvnafpsqps 180
 QY 121 GYTSPTRVHRDSRVSSVSSFTSYQDIYSRARSRSRALHSSDRHNYSPPVNAFPSPSS 180
 Db 181 anssftgcsslgysssrnsrfskasgptqyifheednmfvadaptisrvfaektmynfss 240
 QY 181 ANSSFTGCSLSGYSSSRNSRFSKASGPTQYIFHEEDMNFVDAPTISRVEDEKTMYNFSS 240
 Db 241 prgmclinnnehfegmptrngtkadknltnlfrcmgytvickdnltrgmlltirdfak 300
 QY 241 PRGMCLINNNEHFEGMPTRNKTKADKDLTNLFRCMGYTVICKDNLTRGMLLTIRDFAK 300
 Db 301 heshgdsailvlshgeenviigvddipistheiydlinaanaprlankpkivfvacrg 360
 QY 301 HESHGDSAILVLSHGEENVIIIGVDDIPISTHEIYDLLAANAPRLANKPKIVFVQACRG 360
 Db 361 errdngfpvldsvdvapflrrgwdnrdgplfnflgcvrpqvqvwkpsqadiliry 420
 QY 361 ERDNGFPVLDSVDGVPFLRRGWDNRDGLFNFLGCVRPQVQVWRKPPSQADILIRYA 420
 Db 421 ttaqyvswnsarsgswfigavcefsthvkmdvvelitevnkkyacgftqsgsnlkg 480
 QY 421 TTAQYVSWNSARSGSWFIQAVCEVSTHAKMDVVELLTVNKKYACGFTQSQGSNILKG 480
 Db 481 mpemtsrlkkfyfwearnsav 503
 QY 481 MPEMTSRLKKFYFWPEARNSAV 503

RESULT 4
 ID R53286 standard; Protein: 503 AA.
 AC R53286;
 DT 23-JUN-1994 (first entry)
 DE Ced-3 (A466V).
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain.
 OS Caenorhabditis elegans.
 FH Key location/Qualifiers
 FT misc_difference 466
 FT /label= A466V
 FT /note= "From allele n2430"

WO9325685-A.
 PD 23-DEC-1993.
 PF 14-JUN-1993; U05701.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-979638.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shahan S, Yuan J;
 DR WPI: 94-007542/01.
 DR N-PSDB; Q64743.
 PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 PT develop agents to increase or prevent cell death in organisms
 PS Claim 13; Fig 4; 127pp; English.
 CC The sequences given in R53279-88 are encoded by mutant versions of the
 CC C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3
 CC transcript and was most abundant in embryos, but was also detected in
 CC larvae and young adults, suggesting that ced-3 is expressed not only
 CC in cells undergoing cell death. The four largest introns as well as
 CC sequences 5' of the start codon contain repetitive elements, some of
 CC which have been characterised in non-coding regions of other C. elegans
 CC genes, such as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino
 CC acids in length. Ced-3 is highly hydrophilic with no significant
 CC hydrophobic region that might be a transmembrane region. One region
 CC of Ced-3 is very rich in serine. It is thought that this region is
 CC involved in protein-protein interactions, similar to acid blobs in
 CC transcription factors. Of the mutations which occur within the ced-3
 CC gene, eight of the mutations are missense mutations, two are nonsense
 CC mutations and two are putative splicing mutations. These mutations
 CC establish the null phenotype of the ced-3 gene, confirming that ced-3,
 CC like ced-4, function is not essential for viability. The ced-3 and
 CC ced-4 gene products may be used to develop agents for treating
 CC conditions characterised by cell deaths, such as myocardial infarction,
 CC stroke, degenerative disease, traumatic brain injury, hypoxia,
 CC pathogenic infection, aging or hair loss.
 SQ Sequence 503 AA;

Query Match 99.9%; Score 3599; DB 9; Length 503;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 mmrgdrslrllernimmfshlkvdeilevliakqvlinsdngdmnscgtvrekrrivka 60
 QY 1 MMRODRSLRLERNIMMFSSHLKVEILEVLIKQVLNSDNGDMINSCGTVERKREIVKA 60
 Db 61 vqrrgdvafadafdalrstqheglavleplarsvsnavefecmpspashrralspa 120
 QY 61 VORRGDVAFAFDALRSTGHEGLAEVLEPLARSVSNAVEFECMPSPASHRRALSPA 120
 Db 121 gytsptrvhrdsrvssvssftsdyqdiysrarsrsralhssdrhnysspvnafpsqps 180
 QY 121 GYTSPTRVHRDSRVSSVSSFTSYQDIYSRARSRSRALHSSDRHNYSPPVNAFPSPSS 180
 Db 181 anssftgcsslgysssrnsrfskasgptqyifheednmfvadaptisrvfdektmynfss 240
 QY 181 ANSSFTGCSLSGYSSSRNSRFSKASGPTQYIFHEEDMNFVDAPTISRVEDEKTMYNFSS 240
 Db 241 prgmclinnnehfegmptrngtkadknltnlfrcmgytvickdnltrgmlltirdfak 300
 QY 241 PRGMCLINNNEHFEGMPTRNKTKADKDLTNLFRCMGYTVICKDNLTRGMLLTIRDFAK 300

CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 503 AA;

Query Match 100.0%; Score 3603; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mmrgdrslrllernimmfshlkvdeilevliakqvlnsdngdminsctvrekrrivka 60
1 MMRQDRSLRLERNIMFSSHLKVEILEVLIKQVLNSDNGDMINSCGTVREKREIVKA 60
Cc 1 MMRQDRSLRLERNIMFSSHLKVEILEVLIKQVLNSDNGDMINSCGTVREKREIVKA 60
Db 61 vqrrgdvafdafydlrstgheglaevleplarsvdsnavefecpmaspashrralspa 120
61 VQRRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDSNAVEFECPMSPASHRRALS PA 120
Qy 121 gytsptrvhrdsvssvssftsytgdlysrarsrsrslhssdrhnyssppvnafpsqps 180
121 GYTSPTRVHRDSVSSVSSFTSYQDIYSRARSRSRSLHSSDRHNYSSPPVNAFPSPSS 180
Qy 121 GYTSPTRVHRDSVSSVSSFTSYQDIYSRARSRSRSLHSSDRHNYSSPPVNAFPSPSS 180
Db 181 anssftgcsllysssrnsrfskagptqyifhnedmfvdaptisrvfdektmyrnfs 240
181 ANSFTGCSSLGYSSSRNSRFSKAGPTQYIFHEEDMNEVDAPTISRVEDEKTYRNFS 240
Qy 181 ANSFTGCSSLGYSSSRNSRFSKAGPTQYIFHEEDMNEVDAPTISRVEDEKTYRNFS 240
Db 241 prgmcllnehefegmptnrgtkadkdnltlfrcmgytvickdnltrgmlltrdfak 300
241 PRGMCLLINNEHEFQMPTNRGTAKDKDNLTLFRCMGYTVICKDNLTRGMLLTRDFAK 300
Qy 241 PRGMCLLINNEHEFQMPTNRGTAKDKDNLTLFRCMGYTVICKDNLTRGMLLTRDFAK 300
Db 301 heshgdsalvilshgeenvliigvdipistheiydlinaanaprlankpkivfvacrg 360
301 HESHGDSALVILSHGEENVLIIGVDIPISTHEIYDLINAANAPRLANKPKIVFVACRG 360
Qy 301 HESHGDSALVILSHGEENVLIIGVDIPISTHEIYDLINAANAPRLANKPKIVFVACRG 360
Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnflgcvrpgvqvrwkksqadiliry 420
361 ERRDNGFPVLDSVDGPAPFLRRGWDNRDGPLFNFLGCVRPQVQVWRKKPSQADILIRYA 420
Qy 361 ERRDNGFPVLDSVDGPAPFLRRGWDNRDGPLFNFLGCVRPQVQVWRKKPSQADILIRYA 420
Db 421 ttaqvswrnsargswfigavcevfsthakmdvveltevnkkvacgfgtsqsnllkq 480
421 TTAQVSWRNSARGSWFIGAVCEVFSTHAKMDVVELTEVNKKVACGFGTSQSNLLKQ 480
Qy 421 TTAQVSWRNSARGSWFIGAVCEVFSTHAKMDVVELTEVNKKVACGFGTSQSNLLKQ 480
Db 481 mpemtsrllkkfyfwearnsav 503
481 MPEMTRSLLKKFYFWPEARNSAV 503
Qy 481 MPEMTRSLLKKFYFWPEARNSAV 503

RESULT 2
ID R98754 standard; Protein; 503 AA.
AC R98754;
DT 17-DEC-1996 (first entry)
DE Nematode Ced-3 cell death protein.
KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;
KW cell death; apoptosis; neural degeneration; inflammation;
KW antiinflammatory.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT region 78..205
FT cleavage_site 131 /label= Serine-rich_region
FT cleavage_site 371 /note= "autocleavage site"
FT cleavage_site 371 /note= "autocleavage site"
FT active_site 358 /note= "active site Cysteine residue"
FT active_site 358 /note= "active site Cysteine residue"
PN WO9625946-A1.
PD 29-AUG-1996.
PF 23-FEB-1996; U02473.
PR 24-FEB-1995; US-394189.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 96-425082/42.
DR N-PSDB; T38196.
PT Ced-3 and human interleukin 1-beta convertase genes and proteins

PT useful to treat inflammation and diseases characterised by cell
PT death
PS Claim 18; Fig 3; 139pp; English.
CC The Caenorhabditis elegans Ced-3 protein (R98754) is encoded
CC by the ced-3 cell death gene (T38196). It is structurally similar
CC to human interleukin-1 beta converting enzyme (ICE) (R98755); the
CC 2 proteins have 28% amino acid identity (higher in the C-terminal
CC regions critical for activity), suggesting that Ced-3 may be a
CC cysteine protease like ICE and that ICE may be a human equivalent
CC of the nematode cell death protein. Ced-3 proteins can therefore
CC be used in prodn. of novel drugs for enhancing or inhibiting the
CC activity of ICE for the treatment of inflammatory diseases and/or
CC diseases caused by cell death. Novel inhibitors of ced-3 activity
CC include portions of Ced-3. Ced-3 with a C360A substn. (see also
CC R98758) can prevent programmed cell death in C. elegans.
SQ Sequence 503 AA;

Query Match 100.0%; Score 3603; DB 19; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mmrgdrslrllernimmfshlkvdeilevliakqvlnsdngdminsctvrekrrivka 60
1 MMRQDRSLRLERNIMFSSHLKVEILEVLIKQVLNSDNGDMINSCGTVREKREIVKA 60
Qy 1 MMRQDRSLRLERNIMFSSHLKVEILEVLIKQVLNSDNGDMINSCGTVREKREIVKA 60
Db 61 vqrrgdvafdafydlrstgheglaevleplarsvdsnavefecpmaspashrralspa 120
61 VQRRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDSNAVEFECPMSPASHRRALS PA 120
Qy 61 VQRRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDSNAVEFECPMSPASHRRALS PA 120
Db 121 gytsptrvhrdsvssvssftsytgdlysrarsrsrslhssdrhnyssppvnafpsqps 180
121 GYTSPTRVHRDSVSSVSSFTSYQDIYSRARSRSRSLHSSDRHNYSSPPVNAFPSPSS 180
Qy 121 GYTSPTRVHRDSVSSVSSFTSYQDIYSRARSRSRSLHSSDRHNYSSPPVNAFPSPSS 180
Db 181 anssftgcsllysssrnsrfskagptqyifhnedmfvdaptisrvfdektmyrnfs 240
181 ANSFTGCSSLGYSSSRNSRFSKAGPTQYIFHEEDMNEVDAPTISRVEDEKTYRNFS 240
Qy 181 ANSFTGCSSLGYSSSRNSRFSKAGPTQYIFHEEDMNEVDAPTISRVEDEKTYRNFS 240
Db 241 prgmcllnehefegmptnrgtkadkdnltlfrcmgytvickdnltrgmlltrdfak 300
241 PRGMCLLINNEHEFQMPTNRGTAKDKDNLTLFRCMGYTVICKDNLTRGMLLTRDFAK 300
Qy 241 PRGMCLLINNEHEFQMPTNRGTAKDKDNLTLFRCMGYTVICKDNLTRGMLLTRDFAK 300
Db 301 heshgdsalvilshgeenvliigvdipistheiydlinaanaprlankpkivfvacrg 360
301 HESHGDSALVILSHGEENVLIIGVDIPISTHEIYDLINAANAPRLANKPKIVFVACRG 360
Qy 301 HESHGDSALVILSHGEENVLIIGVDIPISTHEIYDLINAANAPRLANKPKIVFVACRG 360
Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnflgcvrpgvqvrwkksqadiliry 420
361 ERRDNGFPVLDSVDGPAPFLRRGWDNRDGPLFNFLGCVRPQVQVWRKKPSQADILIRYA 420
Qy 361 ERRDNGFPVLDSVDGPAPFLRRGWDNRDGPLFNFLGCVRPQVQVWRKKPSQADILIRYA 420
Db 421 ttaqvswrnsargswfigavcevfsthakmdvveltevnkkvacgfgtsqsnllkq 480
421 TTAQVSWRNSARGSWFIGAVCEVFSTHAKMDVVELTEVNKKVACGFGTSQSNLLKQ 480
Qy 421 TTAQVSWRNSARGSWFIGAVCEVFSTHAKMDVVELTEVNKKVACGFGTSQSNLLKQ 480
Db 481 mpemtsrllkkfyfwearnsav 503
481 MPEMTRSLLKKFYFWPEARNSAV 503
Qy 481 MPEMTRSLLKKFYFWPEARNSAV 503

RESULT 3
ID R53285 standard; Protein; 503 AA.
AC R53285;
DT 23-JUN-1994 (first entry)
DE Ced-3 (A449V).
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophobic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT misc_difference 449 /label= A449V
FT /note= "From allele n1164"
PN WO9325685-A.

3.08; Score 30; DB 16; Length 281;

```
Search completed: Fri Aug 6 04:11:01 1999
Job time : 1735 secs.
```

```
Search completed: Fri Aug 6 04:11:01 1999
Job time : 1735 secs.
```


JOURNAL Unpublished (1998)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797455.

Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyungjido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myunesun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnam@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.

FEATURES
Source Location/Qualifiers

1. 247
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2: XhoI; Directional cDNA library inserted into lambda ZAPII vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1784"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT 7 a 16 c 21 g 34 t 169 others
ORIGIN

Query Match 4.7%; Score 47; DB 17; Length 247;
Best Local Similarity 14.0%; Pred. No. 1.40e-22;
Matches 30; Conservative 97; Mismatches 86; Indels 2; Gaps 2;

18 AMMNKHTHMBBWCVRVGTITNNGKHNGRTTWNDCSDNAHCRTYTBWYARSKYGY 77
1976 AAAAATTCAGAGAGGGTCAAAATTTTCCGATTTTACTGACTTTCACCTTTTTCGTA 2035
78 GTBYSWNVDTNTGCTGVKTTVNVHSGWNNRCSNSVYVWBTAYCDYBHYBDRAHYDD 137
2036 TCAGTGCAGTGTGTG-GAGTTTGTGACGAAAACTAGGAAAAAATCGATAAAATTA 2094
138 TRCTNDRGYCNYTASDNGTSATKRYTGYDKTDSDCGG-CWRKYTYGSSBYBRGVNVMV 196
2095 AAATCGAGCTGAATTTTGAGACAAATTTTAAAAAACAACATAATTTTCCATAATTC 2154
197 RTTSMWTDKSTKMSMDMSRRSRVHYGRMBNKKR 231
2155 ACTCATTTTCAGACTAAATCGAAATCAATCGTA 2189

RESULT 7
LOCUS AA754458 247 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
ACCESSION AA754458
NID 92801164
VERSION AA754458.1 GI:2801164
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE 1 (bases 1 to 247)
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL Unpublished (1998)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797455.
Contact: Eun M.Y.

Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyungjido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myunesun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnam@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.

FEATURES
Source Location/Qualifiers

1. 247
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2: XhoI; Directional cDNA library inserted into lambda ZAPII vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1784"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT 7 a 16 c 21 g 34 t 169 others
ORIGIN

Query Match 4.4%; Score 44; DB 17; Length 247;
Best Local Similarity 11.9%; Pred. No. 2.20e-19;
Matches 29; Conservative 112; Mismatches 99; Indels 3; Gaps 3;

5 TMNTVWRGCCCBAMNKHMTBMBCCVRVGTITNNGKHNGRTTWNDCSDNAHCRTY 64
2799 TCGGCAATCTACAAATATGCGCTTTGGCCACGAAATTAAGTCTTAAAAATAAT 2740
65 TVBMYARSKYGY-GTBYSWNVDTNTGCTGVKTTVNVHSGWNNRCSNSVYVWBTAYC 123
2739 CACTCGAGCTTTCAGTAATTAATTAATTAATTTCACTTCAATTAATTTGGGAAAGTCATAAT 2680
124 DYBHYBDRAHYDDTRCTNDRGYCNYTASDNGTSATKRYTGYDKTDSDCGGCWRKYTYG 183
2679 TTCCGATTTTCGAGGAGAAAAACCGTCCGGTGGTTCGCAAAATTAAGATCCGCCCTTT 2620
184 SSBYBRGVNVMVRTTSMWTDKSTKMSMDMSR-RSRVHYGRMBNKKR-KRGSRRWTDTK 241
2619 TTCCGTGAGACCTTGTCTTTGGCCGAACCTTAAAAACCTACGATCTGCGAGAGGTT 2560
242 TWR 244
2559 CAA 2557

RESULT 8
LOCUS C70143 360 bp mRNA EST 29-SEP-1997
DEFINITION C70143 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA clone yk383f3 5', mRNA sequence.
ACCESSION C70143
NID 92440668
VERSION C70143.1 GI:2440668
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Apr 14, 1993 this sequence version replaced gi:692895.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics

BASE COUNT 109 a 63 c 97 g 91 t
ORIGIN
Query Match 19.8%; Score 198; DB 32; Length 360;
Best Local Similarity 100.0%; Pred. No. 1.95e-213;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 TAAATCGAAATCAATCGTACTCTGACTACGGGTACAGAGAGGTCAACCATACGCCG 61
QY 2169 TAAATCGAAATCAATCGTACTCTGACTACGGGTACAGAGAGGTCAACCATACGCCG 2228

Db 62 AAGATGATCGCTCAAGATAGAGAGAGCTTGTAGAGAGAGCAATTATGATGTTCTTCTAGT 121
QY 2229 AAGATGATCGCTCAAGATAGAGAGAGCTTGTAGAGAGAGCAATTATGATGTTCTTCTAGT 2288

Db 122 CATCTAAAGTCGATGAATTCGAGTTCATCGCAAAACAAGTGTGAATAGTGAT 181
QY 2289 CATCTAAAGTCGATGAATTCGAGTTCATCGCAAAACAAGTGTGAATAGTGAT 2348

Db 182 AATGAGATATGATTAAT 199
QY 2349 AATGAGATATGATTAAT 2366

RESULT 2
LOCUS C39290 360 bp mRNA EST 10-SEP-1997
DEFINITION C39290 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk198f10 5', mRNA sequence.
ACCESSION C39290
NID 92375527
VERSION C39290.1 GI:2375527
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 360)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1393644.

CONTACT: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.

FEATURES
SOURCE location/Qualifiers
1.360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(el489)"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
/db_xref="taxon:6239"
/clone="yk198f10"
/clone_lib="Yuji Kohara unpublished cDNA"

BASE COUNT 102 a 67 c 100 g 91 t
ORIGIN

Query Match 17.7%; Score 177; DB 15; Length 360;
Best Local Similarity 100.0%; Pred. No. 1.76e-185;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 CTCTGACTACGGGTACAGAGAGGTCAACCATCAGCCGAGAGATGATGCGTCAAGATAGA 60
QY 2190 CTCTGACTACGGGTACAGAGAGGTCAACCATCAGCCGAGAGATGATGCGTCAAGATAGA 2249

Db 61 AGGAGCTTGTAGAGAGGACATATGATGTTCTCTAGTCATCTAAAGTCGATGAATTT 120
QY 2317 TTTCTATCGCAAAACAAGTGTGAATAGTGAATATGAGATATGATTAAT 2366

QY 2250 AGGAGCTTGTAGAGAGGACATATGATGTTCTCTAGTCATCTAAAGTCGATGAATTT 2309

Db 121 CTCGAGTTCTCATCGCAAAACAAGTGTGATAGTATGATATGAGATATGATTAAT 177
QY 2310 CTCGAGTTCTCATCGCAAAACAAGTGTGATAGTATGATATGAGATATGATTAAT 2366

RESULT 3
LOCUS C51090 377 bp mRNA EST 11-SEP-1997
DEFINITION C51090 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk491a5 5', mRNA sequence.
ACCESSION C51090
NID 92388343
VERSION C51090.1 GI:2388343
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 377)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1405104.

CONTACT: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.

FEATURES
SOURCE location/Qualifiers
1.377
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(el489)"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
/db_xref="taxon:6239"
/map="11q23"
/clone="yk491a5"
/clone_lib="Yuji Kohara unpublished cDNA"

BASE COUNT 106 a 74 c 103 g 92 t 2 others
ORIGIN

Query Match 16.2%; Score 162; DB 15; Length 377;
Best Local Similarity 98.8%; Pred. No. 1.19e-165;
Matches 168; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 1 TACNGTGCATAGAGAG-TCAACCATCAGCCGAGATGATGCGTCAAGATAGAGAGCT 59
QY 2197 TACGGGTACAGAGAGGTCAACCATCAGCCGAGATGATGCGTCAAGATAGAGAGCT 2256

Db 60 TGCTAGAGAGACATATGATGTTCTCTAGTCATCTAAAGTCGATGAATTTCTCGAAG 119
QY 2257 TGCTAGAGAGACATATGATGTTCTCTAGTCATCTAAAGTCGATGAATTTCTCGAAG 2316

Db 120 TTCTCATCGCAAAACAAGTGTGATAGTGAATATGAGATATGATTAAT 169
QY 2317 TTCTCATCGCAAAACAAGTGTGATAGTGAATATGAGATATGATTAAT 2366

RESULT 4
LOCUS AA754459 252 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPIT cDNA Library Oryza sativa
cDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID 92801165
VERSION AA754459.1 GI:2801165
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa

[M][E][S][E] (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSrch_n n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Fri Aug 6 03:42:06 1999; MasPar time 1720.16 Seconds
1362.179 Million cell updates/sec
Tabular output not generated.

Title: >US-08-287-669-18
Description: (1801-2800) from US08287669.seq (3 of 10)
Perfect Score: 1000
N.A. Sequence: 1801 TTCAGGCTGACAAACAGAA.....CATTTGTAGATTGCCGAA 2800
Comp: AAGTCCGACTGTGTCTT.....GTAAACATCTAACGCTT

Scoring table: TABLE default
Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-est58
1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
6:em_est9 7:em_gss1
Database: genbank-est111
8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
17:gb_est18 18:gb_est19 19:gb_est20 20:gb_est21
21:gb_est22 22:gb_est23 23:gb_est24 24:gb_est25
25:gb_est26 26:gb_est27 27:gb_est28 28:gb_est29
29:gb_est30 30:gb_est31 31:gb_est32 32:gb_est33 33:gb_est34
34:gb_est35 35:gb_est36 36:gb_est37 37:gb_est38 38:gb_est39
39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 11.717; Variance 3.590; scale 3.264

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	198	19.8	360	32	D75533	CELK105H1F Yuji Kohara	1.95e-213
2	177	17.7	360	15	C39290	C39290 Yuji Kohara unp	1.76e-185
3	162	16.2	377	15	C51090	C51090 Yuji Kohara unp	1.19e-165
4	60	6.0	252	17	AA754459	97SN1787 Rice Immature	4.39e-37
5	50	5.0	252	17	AA754459	97SN1787 Rice Immature	7.72e-26
6	47	4.7	247	17	AA754458	97SN1784 Rice Immature	1.40e-22
7	44	4.4	247	17	AA754458	97SN1784 Rice Immature	2.20e-19
8	36	3.6	360	15	C70143	C70143 Yuji Kohara unp	2.95e-11
9	36	3.6	2275	20	AF034173	AF034173 Human mRNA (T	2.95e-11
10	34	3.4	2275	20	AF034173	AF034173 Human mRNA (T	2.46e-09

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
11	31	3.1	580	18	C84753	C84753 Dictyostelium d	1.48e-06
12	31	3.1	779	37	B21852	T2F1-Sp6 TAMU Arabidop	1.48e-06
13	31	3.1	876	37	B12885	T17A8-T7 TAMU Arabidop	1.48e-06
14	30	3.0	281	16	AA633064	ng10e07.s1 NCI_CGAP_Th	1.16e-05
15	30	3.0	401	42	AA045355	GSSC01014 Trypanosoma	1.16e-05
16	30	3.0	503	37	B08074	T7G24-Sp6.1 TAMU Arabi	1.16e-05
17	30	3.0	727	37	B12502	F18F5-Sp6.1 IGF Arabid	1.16e-05
18	30	3.0	809	37	B10155	F4F14-Sp6.1 IGF Arabid	1.16e-05
19	29	2.9	161	24	AI204411	qf72e04.x1 Soares_test	8.77e-05
20	29	2.9	187	8	Z24996	HSB82C032 STRATAGENE H	8.77e-05
21	29	2.9	215	26	AI381412	tc50f01.x1 Soares_tota	8.77e-05
22	29	2.9	231	19	R25172	yh37c06.s1 Soares_plac	8.77e-05
23	29	2.9	244	24	AI198459	qf8a07.x1 Soares_plac	8.77e-05
24	29	2.9	300	21	AA953741	oo37e09.s1 NCI_CGAP_Lu	8.77e-05
25	29	2.9	308	35	AA039740	zf10g10.s1 Soares_feta	8.77e-05
26	29	2.9	323	23	AI147215	qa89c09.x1 Soares_feta	8.77e-05
27	29	2.9	335	8	T18123	0677c3 czapFDD2.1, De	8.77e-05
28	29	2.9	370	15	AA593600	no26d09.s1 NCI_CGAP_Pr	8.77e-05
29	29	2.9	374	18	AA783005	ai58a06.s1 Soares_test	8.77e-05
30	29	2.9	374	20	AA918212	on82g01.s1 Soares_NFL	8.77e-05
31	29	2.9	377	30	R62771	yi10d10.s1 Soares_plac	8.77e-05
32	29	2.9	388	24	AI220889	qg03c07.x1 Soares_plac	8.77e-05
33	29	2.9	424	16	AA622274	no42h08.s1 NCI_CGAP_Pr	8.77e-05
34	29	2.9	440	20	AA907140	od44c05.s1 NCI_CGAP_GC	8.77e-05
35	29	2.9	441	23	AI138427	qd84d07.x1 Soares_test	8.77e-05
36	29	2.9	444	23	AI138414	qd84c04.x1 Soares_test	8.77e-05
37	29	2.9	456	32	N26720	yx61h09.s1 Soares_mela	8.77e-05
38	29	2.9	494	35	AA039776	zf10c10.s1 Soares_feta	8.77e-05
39	29	2.9	505	29	AI568541	tn40e10.x1 NCI_CGAP_Br	8.77e-05
40	29	2.9	505	17	AA722943	zg85a06.s1 Soares_feta	8.77e-05
41	29	2.9	525	24	AI188773	qd16f05.x1 Soares_plac	8.77e-05
42	29	2.9	535	35	AA045000	zk63g04.s1 Soares_preg	8.77e-05
43	29	2.9	541	30	H09727	yl190e11.s1 Soares_infa	8.77e-05
44	29	2.9	601	15	AA573749	nk07b07.s1 NCI_CGAP_Co	8.77e-05
45	29	2.9	844	37	B10796	T26G15-Sp6 TAMU Arabid	8.77e-05

ALIGNMENTS

RESULT 1
LOCUS D75533 360 bp mRNA
DEFINITION CELK105H1F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk105h1 5', mRNA sequence.
ACCESSION D75533
NID G1121320
VERSION D75533.1 GI:1121320
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 360)
AUTHORS Kohara, Y., Mitsuki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and
Tabara, H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT

FEATURES
source
1..360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(el489)"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
/db_xref="taxon:6239"
/clone="yk105h1"

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykoha@ddbj.nig.ac.jp
Insert Length: 2279 Std Error: 0.00.
Location/Qualifiers

FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT repeat_unit 5221..5330
FT /*tag= q
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon 5547..5760
FT /*tag= r
FT /number= Exon_5
FT intron 5761..5814
FT /*tag= s
FT /number= Intron_5
FT exon 5815..5942
FT /*tag= t
FT /number= Exon_6
FT intron 5943..6297
FT /*tag= u
FT /number= Intron_6
FT repeat_region 6062..6138
FT /*tag= v
FT /rpt_type= INVERTED
FT exon 6298..6537
FT /*tag= w
FT /number= Exon_7
FT intron 6538..7012
FT /*tag= x
FT /number= Intron_7
FT repeat_unit 6567..6625
FT /*tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /*tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7012..7075
FT /*tag= aa
FT /number= Exon_8
PD WO9325685-A.
PD 23-DEC-1993.
PE 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PI (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.
DR P-PSDB; R53280.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PS develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127pp: English.
CC The sequences given in Q64735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

Query Match 98.8%; Score 988; DB 9; Length 7653;
Best Local Similarity 99.4%; Pred. No. 0.00e+00;
Matches 994; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Db 1801 ttccagctgacaaacagaaacaaacacacacacacattttaaaatcagtttccaat 1860
QY 1801 ttccagcctgacaaacagaaacaaacacacacacacacattttaaaatcagtttccaat 1860
Db 1861 taaaaataacgatttcattgaaaattgtgtttatgtttgccaaaataaagagaact 1920
QY 1861 TAAAAATAACGATTTCATTGAAAAATTGTTTATGTTTGGCAAAATAAAGAGAACT 1920
Db 1921 gattcaaaaacaattttaaaaaaaaccccaaaattcgccaagaatcaagataaaaa 1980
QY 1921 GATTCAAAAACAATTTTAACAAAAAACCCCAAAATTCGCCAGAAATCAAGATAAAAA 1980
Db 1981 ttcaagagggtcaaaaatttcgatttactgacttccacttttttcgtagtcagt 2040
QY 1981 TTCAAGAGGGTCAAAAATTTCCGATTTCGATTTCACCTTTTTCGTAGTTCAGT 2040
Db 2041 gcagttgttgagtttttgacgaaactagaaanaaatcgataaaaattactcaaatcg 2100
QY 2041 GCAGTTGTTGAGTTTTCGAGAAACTAGGAAAAAATCGATAAAAATTTACTCAAAATCG 2100
Db 2101 agctgaattttgagacaaatgtttaaaaaaaacactattttccaaataattcactcat 2160
QY 2101 AGCTGAATTTTGAGGACAAATGTTAAAAAAAACACTATTTTCCAAATAATTTCACTCAT 2160
Db 2161 ttccagactaaatcgaaaaatcaaatcgtactctgactaagctcagtagaagagtcaccc 2220
QY 2161 TTTCAGACTAAATCGAAAAATCAAAATCGTACTCTGACTACGGGTCAGTAGAGAGTCAACC 2220
Db 2221 atcagccgaagatgagtcgtaagaatagagagcttgtagagagagacattatgagt 2280
QY 2221 ATCAGCCGAAGATGATGCGTCAAGATAGAAAGAGAGCTTGCTAGAGAGACATTATGATGT 2280
Db 2281 ttctctagtcacttaaaagtcgatgaatttctcgaagttctcatcgcaaaacagtggtga 2340
QY 2281 TCTCTAGTCATCTAAAGTCGATGAATTTCTCGAGTTCTCATCGCAAAACAAGTGTGA 2340
Db 2341 atagtgataatgagagatgatgaattatgtgagttttaatcgaaataaattttaaaaaa 2400
QY 2341 ATAGTGATAATGAGATGATGATTAATGTGAGTTTAAATCGAATAATAATTTAAAAAAA 2400
Db 2401 aattgataataataaagaataattttgcagtcagtgtgaaacggttcgcgagaagaagacggg 2460
QY 2401 AATTGATATATATAAGAAATATTTTTCGAGTCATGTGAAACGGTTCGCGAAGAGACGGG 2460
Db 2461 agatcgtgaaagcagtcgaaacgagagagatgtggttcacacggtttatgatgtc 2520
QY 2461 AGATCGTGAAGCAGTGCACACGAGGAGATGTGGCGTTTCGACGCGTTTATGATGCTC 2520
Db 2521 ttgcctctacgagcagcgaagactgtcgaagttctgaagctctgcgcagatcgtag 2580
QY 2521 TTGCGCTACGGGACACGAGGACTTGCTGAAGTTCTTGAACCTCTCGCAGATCGTAGG 2580
Db 2581 ttttaagttcggcgcaaaagcaaggtctcaagyaaaaaaagagcgcatcgtattt 2640
QY 2581 TTTTAAAGTTCGGCGCAAAAGCAAGGCTCTCACGGAAAAAGAGGCGGATCGTATTTT 2640
Db 2641 gcaacccacggcagcgttttttctcctcgaaaatcggaaattatgcactttcccaatat 2700
QY 2641 GCAACCCACGGCAGCGGCTTTTCTCTCGAAAAATCGGAAATTATGCACTTTCCCAATAT 2700
Db 2701 ttgaagtgaaatataatttactgaagagtcgaagtgatattttttaacacta 2760
QY 2701 TTGAAGTGAATATATATTTTACTGAAAGCTCGAGTGATATTATTTTAAACACTA 2760
Db 2761 atttcgtgagcgaagacattttagatttgcgaa 2800
QY 2761 ATTTTCGTGGCGCAAAAGCCATTTTGTAGATTGCGCGAA 2800


```
OY 2521 TTGCTCTACGGGACACGAGACTTGTGAAGTCTTGAACCTCTCGCCAGATCGTAGG 2580
Db 2581 ttttaagttcggcgcaaaagcaagggctcaccgaaaaaagagccgacgtatttt 2640
OY 2581 TTTTAAAGTTTCGCCCAAAAGCAAGGCTCTCACGGAAAAAGAGCGGATCGTAATTT 2640
Db 2641 gcaacccaccgacggttttttccctccgaaaaatcggaaattatgcactttcccaatat 2700
OY 2641 GCAACCCACCACGCGCGGTTTTCCTCCGAAAAATCGGAAATATATGCACCTTCCCAATAT 2700
Db 2701 ttgaagtgaataataatttattactgaaaagctcgagtgattattatttttaacacta 2760
OY 2701 TTGAAGTGAATAATATTTTATTACTGAAAGCTCGAGTGATTTATTATTTTTAACACTA 2760
Db 2761 atttcgtggcgcaaaagccattttagatttgcgaa 2800
OY 2761 ATTTTCGTGCGCGCAAAAGCCATTTTGTAGATTGCGCGAA 2800

RESULT 12
ID 064739 standard; DNA; 7653 BP.
AC 064739;
DT 23-JUN-1994 (first entry)
DE ced-3 (G6297A) gene.
KW C. elegans; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key location/Qualifiers
FT repeat_unit 1356..1472
FT FT /*tag= a
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t repeat at 1490-1614"
FT FT 1490..1614
FT FT /*tag= b
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT FT 2232..2366
FT FT /*tag= c
FT FT /number= Exon_1
FT FT 2367..2429
FT FT /*tag= d
FT FT /number= Intron_1
FT FT 2450..2575
FT FT /*tag= e
FT FT /number= Exon_2
FT FT 2576..2853
FT FT /*tag= f
FT FT /number= Intron_2
FT FT 2854..3107
FT FT /*tag= g
FT FT /number= Exon_3
FT FT 3108..4302
FT FT /*tag= h
FT FT /number= Intron_3
FT FT 3126..3243
FT FT /*tag= i
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT FT 3329..3396
FT FT /*tag= j
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT FT 3487..3759
FT FT /*tag= k
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT FT 3782..4070
FT FT /*tag= l
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 3487-3759"
FT FT 4303..4634
FT FT exon
```

```
FT FT /*tag= m
FT FT /number= Exon_4
FT FT 4635..5546
FT FT /*tag= n
FT FT /number= Intron_4
FT FT 4688..4719
FT FT /*tag= o
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5221..5330
FT FT /*tag= p
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5547..5760
FT FT /*tag= q
FT FT /number= Exon_5
FT FT 5761..5814
FT FT /*tag= r
FT FT /number= Intron_5
FT FT 5815..5942
FT FT /*tag= s
FT FT /number= Exon_6
FT FT 5943..6297
FT FT /*tag= t
FT FT /number= Intron_6
FT FT 5757
FT FT /*tag= u
FT FT /note= "C>T, from allele n717"
FT FT 6062..6138
FT FT /*tag= v
FT FT /rpt_type= INVERTED
FT FT 6298..6537
FT FT /*tag= w
FT FT /number= Exon_7
FT FT 6538..7012
FT FT /*tag= x
FT FT /number= Intron_7
FT FT 6567..6625
FT FT /*tag= y
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT FT 6905..6965
FT FT /*tag= z
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 6567-6625"
FT FT 7012..7075
FT FT /*tag= aa
FT FT /number= Exon_8
FT FT 709325685-A.
FT FT 23-DEC-1993.
FT FT PD 14-JUN-1993; U05701.
FT FT PF 12-JUN-1992; US-897788.
FT FT PR 20-NOV-1992; US-979638.
FT FT PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT FT PI Horvitz HR, Shaham S, Yuan J;
FT FT DR WPI; 94-007542/01.
FT FT PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT FT develop agents to increase or prevent cell death in organisms
FT FT PS Claim 14; Fig 4; 127pp; English.
FT FT CC The sequences given in Q64735-45 represent mutations of the C. elegans
FT FT CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT FT CC was most abundant in embryos, but was also detected in larvae and young
FT FT CC adults, suggesting that ced-3 is expressed not only in cells undergoing
FT FT CC cell death. The four largest introns as well as sequences 5' of the
FT FT CC start codon contain repetitive elements, some of which have been
FT FT CC characterised in non-coding regions of other C. elegans genes, such
FT FT CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
FT FT CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
FT FT CC region that might be a transmembrane region. One region is involved in
FT FT CC very rich in serine. It is thought that this region is involved in
FT FT CC protein-protein interactions, similar to acid blobs in transcription
FT FT CC factors. Of the mutations which occur within the ced-3 gene, eight of
FT FT CC the mutations are missense mutations, two are nonsense mutations and
```



```
FT repeat_unit 3487..3759
FT /tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT repeat_unit 3782..4070
FT /tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FI exon 4303..4634
FI /tag= m
FI /number= Exon_4
FI intron 4635..5546
FI /tag= n
FI /number= Intron_4
FI repeat_unit 4688..4719
FI /tag= o
FI /rpt_type= INVERTED
FI /note= "Repeat 3"
FT repeat_unit 5221..5330
FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon 5547..5760
FT /tag= q
FT /number= Exon_5
FT intron 5761..5814
FT /tag= r
FT /number= Intron_5
FT exon 5815..5942
FT /tag= s
FT /number= Exon_6
FT intron 5943..6297
FT /tag= t
FT /number= Intron_6
FT repeat_region 6062..6138
FT /tag= u
FT /rpt_type= INVERTED
FT exon 6298..6537
FT /tag= v
FT /number= Exon_7
FT intron 6538..7012
FT /tag= w
FT /number= Intron_7
FT repeat_unit 6567..6625
FT /tag= x
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7012..7075
FT /tag= z
FT /number= Exon_8
FT mutation 7020
FT /tag= w
FT /note= "C>T, from allele n1163"
PN W09325685-A.
PD 23-DEC-1993.
PE 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MAST ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shahan S, Yuan J;
DR WPI; 94-007542/01.
P-PSDB; R53288.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127pp; English.
CC The sequences given in Q64735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
```

```
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;
```

```
Query Match 99.0%; Score 990; DB 9; Length 7653;
Best Local Similarity 99.5%; Pred. No. 0.00e+00;
Matches 995; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
Db 1801 ttccagcgtgacaaacagaacaaacaaacaaacattttaaaatcagtttcaaat 1860
QY 1801 TTCAGGCGTGACAAACAGAAACAAACAAACAAACATTTTAAATCAGTTTCAAA 1860
Db 1861 taaaaataacgatttcctcattgaaatgtgtttatgtttgccaaataaagaagact 1920
QY 1861 TAAATAATACGATTCTCATTTGAAAAATGTGTTTATGTTGCCAAAAATAAAGAGAACT 1920
Db 1921 gattcaaacacattttaacaaacaaacaaacaaacatttcgcagaatacaagataaaaa 1980
QY 1921 GATTCAAAACAAATTTTAAACAAAAAAACCCCAAAATTCCGCCAGAAATCAAGATTAATAA 1980
Db 1981 ttcaagagggcacaattttccgatttactgacttccacttttttcgtagttcagt 2040
QY 1981 TTCAAGAGGGTCAAAATTTTCCGATTTTACTGACTTTACCTTTTTCGTAGTTCAGT 2040
Db 2041 gcagttgttggagtttttgacgaaactaggaacaaataatcgataaaataactcaaatcg 2100
QY 2041 GCAGTTGTGGAGTTTTCGACGAAACTAGGAAAAAAATTCGATAAAATTTACTCAAAATCG 2100
Db 2101 agctgaattttgaggacaatgttttaaaaaaaacactattttccaataattcactcat 2160
QY 2101 AGCTGAATTTTGAGGACAATGTTTAAAAAAAACACTTATTTTCCAAATATTCACATCAT 2160
Db 2161 ttccagactaaatcgaaatcaaatcgtactctgactacggtcagtagagagtcaccc 2220
QY 2161 TTTCAGACTAAATCGAAATCAATCGTACTCTGACTACGGGTCAGTAGAGAGGTCACACC 2220
Db 2221 atcagccgaagatgatgctgcataagatagaagagcttgctagagaggaacatatgtat 2280
QY 2221 ATCAGCCGAAGATGATGCGTCAGATAGAAAGAGAGCTTGCTAGAGAGAACATATGATGT 2280
Db 2281 tctctagtcacttaaaagtcgatgaatctcgaagttctcatcgcacaaacaaagtgtga 2340
QY 2281 TCTCTAGTCATCTAAAGTCGATGAATTTCTCGAAGTTCATCGCAAAACAGTGTGA 2340
Db 2341 atagtgataatgagagatgatgaatgtgagtttttaacgataataatttaaaaaaa 2400
QY 2341 ATAGTGATAATGAGAGATGATGAATTAATGTGAGTTTAAATCGAATAATATTTTAAAAAAA 2400
Db 2401 aattgataataataagaatatattttgcaagtcacgtggaacggttcgcgagaagagacggg 2460
QY 2401 AATTGATAATATAAAGATAATTTTTCAGATCATGTGGAACGGTTCCGCAAGAGACGGG 2460
Db 2461 agatcgtgaagcagtcacaacgacgagggagagatggttcacacgctttatgatgtc 2520
QY 2461 AGATCGTGAAGCAGTCACACGACGGGAGAGATGGCGTTTCGACGCGTTTATGATGCTC 2520
Db 2521 ttgcgtctacggacaggaagactgtgtaagttcttgaagctctgcacagatcgtag 2580
QY 2521 ttgcgtctacggacaggaagactgtgtaagttcttgaagctctgcacagatcgtag 2580
```


Db 2281 tctctagtcattcaaaagtcgatgaaattctcgaagttctcatcgcaaaaacaagtgtga 2340
|||||
QY 2281 TCTCTAGTCATCAATAAGTCGATGAAATCTCGAAGTCTCATCGCAAAAACAAGTGTGA 2340
Db 2341 atagtgataatgagatatgatatagtgagttttaatcgataataaatttaaaaaa 2400
|||||
QY 2341 ATAGTGATAATGAGATATGATTATGTGAGTTTAAATCGAATAATAATTTAAAAAAA 2400
Db 2401 aattgataataaagaatatttttcgagtcagtgtggaacggttcgcgagaagagacgg 2460
|||||
QY 2401 AATTGATAATAAAGATATTTTTCAGTCATGTGAACGGTTCGCAGAGAAGACGGG 2460
Db 2461 agatcgtgaaagcagtcgaacgagggagatgtggcgttcacagcgttttatgctc 2520
|||||
QY 2461 AGATCGTGAAGCAGTCCACGACGGGGAGATGTGGCGTTCGACCGGTTTATGATGCTC 2520
Db 2521 ttgcctctacggagacgaaggacttgctgaagttcttgaagctctcgccagatcgtagg 2580
|||||
QY 2521 TTCGCTCTACGGGACGACGAGGACTTGCTGAAGTCTTGAACCTCTCGCCAGATCGTAGG 2580
Db 2581 tttttaagtcggcgcaaaagcaagggtctcagcgaaaaagagccgacgtlaattt 2640
|||||
QY 2581 TTTTAAAGTTCGGCGCAAAACCAAGGCTCTCAAGGAAAAAGAGCGGATCGTAATTTT 2640
Db 2641 gcaaccaccgagcggttttctccgaaatccgaaatattatgcacttcccaatat 2700
|||||
QY 2641 GCAACCCACCGGACGGTTTTTCTCCCGAAATCGGAATATGACATTTCCCAATAT 2700
Db 2701 ttgaagtgaatatatttttactgaagctcgagtgattatttttaacacta 2760
|||||
QY 2701 TTGAAGTGAATATATTTTACTGAAGCTCGAGTGATTTATTTTAACACTA 2760
Db 2761 atttcgtggcgcaaaagccattttagatttgcgaa 2800
|||||
QY 2761 ATTTTCGTGGCGCAAAAGCCATTTGTAGATTGCCGAA 2800

RESULT 10

TD 054737 standard; DNA; 7653 BP.

AC 064737;

DE 23-JUN-1994 (first entry)

KW ced-3 (G5757A) gene.

KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;

KW embryogenesis; cell death; hydrophilic; transmembrane; region;

KW hydrophobic; mutation; amino acid; substitution; RNA splicing;

KW protein synthesis; null phenotype; calcium-binding domain; ss.

OS Caenorhabditis elegans.

FH Key Location/Qualifiers

FT repeat_unit 1356..1472

FT /*tag= a

FT /rpt_type= INVERTED

FT /note= "Inverted w.r.t repeat at 1490-1614"

FT repeat_unit 1490..1614

FT /*tag= b

FT /rpt_type= INVERTED

FT /note= "Inverted w.r.t. repeat at 1356-1472"

FT repeat_unit 2232..2366

FT /*tag= c

FT /number= Exon_1

FT repeat_unit 2367..2429

FT /*tag= d

FT /number= Intron_1

FT repeat_unit 2450..2575

FT /*tag= e

FT /number= Exon_2

FT repeat_unit 2576..2853

FT /*tag= f

FT /number= Intron_2

FT repeat_unit 2854..3107

FT /*tag= g

FT /number= Exon_3

FT repeat_unit 3108..4302

FT /*tag= h

FT /number= Intron_3
FT repeat_unit 3126..3243
FT /*tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit 3329..3396
FT /*tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT repeat_unit 3487..3759
FT /*tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT repeat_unit 3782..4070
FT /*tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT exon 4303..4634
FT /*tag= m
FT /number= Exon_4
FT repeat_unit 4635..5546
FT /*tag= n
FT /number= Intron_4
FT repeat_unit 4688..4719
FT /*tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT repeat_unit 5221..5330
FT /*tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon 5547..5760
FT /*tag= q
FT /number= Exon_5
FT mutation 5757
FT /*tag= r
FT /note= "G>A, from allele n2433"
FT repeat_unit 5761..5814
FT /*tag= s
FT /number= Intron_5
FT repeat_unit 5815..5942
FT /*tag= t
FT /number= Exon_6
FT repeat_unit 5943..6297
FT /*tag= u
FT /number= Intron_6
FT repeat_unit 6062..6138
FT /*tag= v
FT /rpt_type= INVERTED
FT repeat_unit 6298..6537
FT /*tag= w
FT /number= Exon_7
FT repeat_unit 6538..7012
FT /*tag= x
FT /number= Intron_7
FT repeat_unit 6567..6625
FT /*tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /*tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT repeat_unit 7012..7075
FT /*tag= aa
FT /number= Exon_8
PN WO9325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
Horvitz HR, Shaham S, Yuan J;


```
FT /number= Exon_2
FT 2576..2853
FT /*tag= f
FT /number= Intron_2
FT 2854..3107
FT /*tag= g
FT /number= Exon_3
FT 3108..4302
FT /*tag= h
FT /number= Intron_3
FT 3126..3243
FT /*tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT 3329..3396
FT /*tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT 3487..3759
FT /*tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT 3782..4070
FT /*tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT 4303..4634
FT /*tag= m
FT /number= Exon_4
FT 4635..5546
FT /*tag= n
FT /number= Intron_4
FT 4688..4719
FT /*tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5221..5330
FT /*tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5547..5760
FT /*tag= q
FT /number= Exon_5
FT 5761..5814
FT /*tag= r
FT /number= Intron_5
FT 5815..5942
FT /*tag= s
FT /number= Exon_6
FT 5943..6297
FT /*tag= t
FT /number= Intron_6
FT 6062..6138
FT /*tag= u
FT /rpt_type= INVERTED
FT 6298..6537
FT /*tag= v
FT /number= Exon_7
FT 6322
FT /*tag= w
FT /note= "C>T, from allele n1949"
FT 6538..7012
FT /*tag= x
FT /number= Intron_7
FT 6567..6625
FT /*tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT 6905..6965
FT /*tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT 7012..7075
FT exon
```

```
FT /*tag= aa
FT /number= Exon_8
FT WO9325685-A.
PN 23-DEC-1993.
PD 14-JUN-1993; U05701.
PF 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shahan S, Yuan J;
DR WPI; 94-007542/01.
DR P-PSDB; R53283.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127pp; English.
CC The sequences given in Q64735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as Fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.0%; Score 990; DB 9; Length 7653;
Best Local Similarity 99.5%; Pred. No. 0.00e+00;
Matches 995; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1801 ttccagctgacaacaagaacaacaacaacacattttaaaatcagttttcaat 1860
QY 1801 TTCCAGGCTGACAACAAGAACAACAACAACAACATTTTAAAAATCAGTTTCAAT 1860
Db 1861 taaaataacgatttctcattgaaatgtgttttatgtttgcccataaagaagaact 1920
QY 1861 TAAAAATAACGATTTCTCATTTGAAAAATGTGTTTATGTTCGCAAAATAAAAAGAACT 1920
Db 1921 gattcaaaaacaattttaacaaaaaaaccccaaaatcgccagaataatcaagataaaaa 1980
QY 1921 GATTCAAAACATTTTAAACAAAAAAACCCCAAAATTCGCCAGAAATCAAGATAAAAAA 1980
Db 1981 ttcaagaggtcaaaaattttccgattttactgactttcaccttttttcgtagttcagt 2040
QY 1981 TTCAAGAGGGTCAAAATTTTCCGATTTTACTGACTTTCACCTTTTTCGTAGTTCAGT 2040
Db 2041 gcagttgttgagatttttgacgaaaaactagaaaaaaatcgataaaaaattactcaatcg 2100
QY 2041 GCAGTTGTTGAGTTTTGTGACGAAAACTAGGAAAAAAATCGATAAAAAATTACTCAAAATCG 2100
Db 2101 agctgaatttgagagacaatgttttaaaaaaaacacactattttccaataattcactcat 2160
QY 2101 AGCTGAATTTTGAGACAATGTTTAAAAAAAACACTATTTTCCAATAATTCTACTCAT 2160
Db 2161 ttccagactaaatcgaaaatcaaatcgtactctgactacggtcagtagagaggtcaacc 2220
QY 2161 TTTCAGACTAAATCGAAATCAAAATCGTACTCTGACTACGGGTCAGTAGAGAGGTCAACC 2220
Db 2221 atcagccgaagaatgatgctcaagatagaaaggagctgtcgtagagaaagacattatgatgt 2280
QY 2221 ATCAGCCGAGAGATGATGCGTCAAGATAGAAAGAGAGCTTGCTAGAGAGGAACATTATGATGT 2280
```



```
FT repeat_unit 6567..6625
FT /tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7012..7075
FT /tag= aa
FT /number= Exon_8
PN W09325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.
DR P-PSDB; R53286.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127p; English.
CC The sequences given in 064735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.0%; Score 990; DB 9; Length 7653;
Best Local Similarity 99.5%; Pred. No. 0.00e+00;
Matches 995; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1801 ttccagctgacaaacagaacaaacacacacacatttaaaatcagtttcaaat 1860
QY 1801 TTCAGGCTGACAAACAGAAACAAAACACACAAACATTTTAAAAATCAGTTTCAAAAT 1860
Db 1861 taaaataacgattctcatatgaaatgttttatgtttgccaataaaagaagact 1920
QY 1861 TAAAAATAACGATTTCTCATTTGAAATGTGTTTATGTTTGGAAATTAAGAGAACT 1920
Db 1921 gattcaaacacattttaacaaacaaaccccaaaatcgccagaatcaagataaaaaa 1980
QY 1921 GATTCAAAACAATTTTAACAAAAAACCCCAAAATTCGCCAGAAATCAAGATAAAAAA 1980
Db 1981 ttcaagagggtcaaaattttccgatttctactacttccaccttttttcgtagttcagt 2040
QY 1981 TTCAAGAGGGTCAAAATTTTCCGATTTTACTGACTTTCACCTTTTTCGTAGTTCAGT 2040
Db 2041 gcaagtgttgagatttttgaacgaacaaactaggaacaaacatcgataaaatctcaaatcg 2100
QY 2041 GCAAGTGTGAGATTTTGACGAAACACTAGGAAAAAATCGATAAAAATTACTCAAAATCG 2100
Db 2101 agctgaattttgagacaaatgtttaaaaaaacactattttccaataattcactcat 2160
QY 2101 AGCTGAATTTTGAGCAACATGTTTAAAAAAAACACTATTTTCCAAATAATTCACTCAT 2160
```

```
Db 2161 ttccagactaaatcgaaaatcaaatcgctactctgactacggtcagtagagagtcacac 2220
QY 2161 TTCCAGACTAAATCGAAAATCAAAATCGTACTCTGACTACGGGCTCAGTAGAGAGGTCAACC 2220
Db 2221 atcagccgaagatgatgcgtcaagatagaaggagctgtctagaaggaacattatgatgt 2280
QY 2221 ATCAGCCGAAGATGATGCGTCAAGATAGAGAGAGCTTGTCTAGAGAGGAACATTATGATGT 2280
Db 2281 tctctagtcattcaaaagtcgatgaatcttcgaagttctcatcgcaaaacaagtgtga 2340
QY 2281 TCTCTAGTCATCTAAAGTCGATGAATTTCTCGAAGTTCTCATCGCAAAAACAAGTGTGA 2340
Db 2341 atagtataatgagatatgatattatgtgagttttaatcgaataataatttaaaaaaa 2400
QY 2341 ATAGTATAATGAGATATGATTAATGTGAGTTTAAATCGAATAATAATTTAAAAAAA 2400
Db 2401 aattgataataaagaataattttgcagtcagtgtgaacggttcgcgagaagagacggg 2460
QY 2401 AATTGATAATATAACAAATATTTTTCAGTCATGTGGAACGGTTCGCCGAGAAGACGGG 2460
Db 2461 agatcgtgaaagcagtgcaacgacgaggagatgtgagttccacggttttatgctc 2520
QY 2461 AGATCGTGAAGCAGTGCACACGACGGGAGATGTGGCGTTTCGACCGCTTTATGATGCTC 2520
Db 2521 ttgcctctacgggacacgaagagactgtgaagttcttgaagcttcgccagatcgtag 2580
QY 2521 TTCGCTCTACGGGACACGAAGGACTTGTGAAGTCTTGAACCTCTGCCAGATCGTAGG 2580
Db 2581 ttttaaaagtcggcgcaaaagcaagaggtctcaacgaaacaaagagccgactcgtaat 2640
QY 2581 TTTTAAAGTTCGGCGCAAAAGCAAGGCTCTCACGAAAAAAGAGCGGATCGTAATTTT 2640
Db 2641 gcaaccaccgagcaggttttccctccgaaaatcggaaatlatgacatttcccaatat 2700
QY 2641 GCAACCCACCGCAGGTTTTTCTCCGAAAATCGGAATATATGCACTTTCCCAATAT 2700
Db 2701 ttgaagtgaaatatatttatttactgaaagctcgagtgatattatttttaacacta 2760
QY 2701 TTGAAGTGAATATATTTTATTACTGAAGCTCGAGTGATTTATTTTAAACACTA 2760
Db 2761 atttcgtgcgcaaaagggccatttctgtagatttgcgaa 2800
QY 2761 ATTTTCGTGGCGCAAAAGGCCATTTGTAGATTGCCGAA 2800

RESULT 9
ID Q64740 standard; DNA; 7653 BP.
AC Q64740;
DT 23-JUN-1994 (first entry)
DE ced-3 (C6322T) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key location/Qualifiers
FT repeat_unit 1356..1472
FT /tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT repeat_unit 1490..1614
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT /tag= c
FT /number= Exon_1
FT intron 2367..2429
FT /tag= d
FT /number= Intron_1
FT exon 2450..2575
FT /tag= e
```


OY 1981 TTCAAGAGGGTCAAAATTTCCGATTTTACTGACTTTCACCTTTTTTTCGTACTCAGT 2040
Db 2041 gcagttgttgaggtttttgacgcaaaactagcaaaaaatcgaataaattactcaaatcg 2100
OY 2041 GCAGTTGTTGGAGTTTTCGACGAAAC TAGGAAAAAATCGATAAAATTTACTCAAAATCG 2100
Db 2101 agctgaattttgaggaacaatgtttaaaaaaaacactattttcccaataattcaactcat 2160
OY 2101 AGCTGAATTTTGAGGACAAATGTTAAAAAACAATAATTTTCCAAATAATTTCACTCAT 2160
Db 2161 ttccagactaaatcgaaaatcgaatcgtactctgactacgctcagtagagaggtcaacc 2220
OY 2161 TTTTCAGACTAAATCGAAAAATCAAAATCGTACTCTGACTACGGGTCAGTAGAGAGTCAACC 2220
Db 2221 atcagccgaagatgatgctcaagatagaagagcttgtagagaggaacatatgatgt 2280
OY 2221 ATCAGCCGAAGATGATGCGTCAAGATAGAAGAGCTTGCTAGAGAGAACATTAATGATGT 2280
Db 2281 tctctagtcataaagaatcgatgaatctcgaagttctcatcgcaaaacaagtgttga 2340
OY 2281 TCTCTAGTCATCTAAAGTCGATGAATTTCTCGAAGTTCTCATCGCAAAACAAGTGTGA 2340
Db 2341 atagtataatgagatatgataatgtgagtttttaatcgaataataattttaaaaaa 2400
OY 2341 ATAGTGATTAATGAGATATGATTAATGTGAGTTTAAATCGAATAATAATTTAAAAAAA 2400
Db 2401 aattgataataaagaataatttttcagtcatgttgaaacggttcgcgagaagaagacggg 2460
OY 2401 AATTGATTAATATAAAGAAATATTTTTCAGTCAATGTGAACGGTTTCGCGAGAAGACGGG 2460
Db 2461 agatcgtgaaagcagtgcaacgacgggagagatgtgacgttcacacgctttatgatgtc 2520
OY 2461 AGATCGTGAAGCAGTGCAACGACGGGAGATGTGGCGTTTCGACGGCTTTATGATGCTC 2520
Db 2521 ttgcctctacgggacacgaagagacttgctgaagttcttgaagcttcgccagatcgtagg 2580
OY 2521 TTGCTCTACGGGACACGAAGACTTGCTGAAGTCTTGAACCTCTCGCCAGATCGTAGG 2580
Db 2581 ttttaaaagtctggcgcaaaagcaagggtctcaccggaaaaaaagagccgatcgtaattt 2640
OY 2581 TTTTAAAGTTGCGCGCAAAAGCAAGGCTCTCACGGAAAAAAGAGCGGATCGTAATTT 2640
Db 2641 gcaacccacggcgacggtttttctcccgaaaatcggaaattatgcactttcccaaatat 2700
OY 2641 GCAACCCACCGGACGCGTTTCTCTCCGAAATCGGAAATTATGCACCTTCCCAATAT 2700
Db 2701 ttgaagtgaataataattattactgaaaagctcgagtgattatttttaaacacta 2760
OY 2701 TTGAAGTGAATATATTTTATTACTGAAAGCTCGAGTGATTTATTTTAAACACTA 2760
Db 2761 atttcgtggcgcaaaagccatttgtagatttgccgaa 2800
OY 2761 ATTTTCGTGGCGCAAAAGCCCATTTGTAGATTGCCGAA 2800

FT FT exon /note= "Inverted w.r.t. repeat at 1356-1472"
FT FT exon /tag= c
FT FT intron /number= Exon_1
FT FT intron /tag= d
FT FT exon /number= Intron_1
FT FT exon /tag= e
FT FT intron /number= Exon_2
FT FT intron /tag= f
FT FT exon /number= Intron_2
FT FT exon /tag= g
FT FT intron /number= Exon_3
FT FT intron /tag= h
FT FT repeat_unit /number= Intron_3
FT FT repeat_unit /tag= i
FT FT repeat_unit /note= "Inverted w.r.t. repeat at 3329-3396"
FT FT repeat_unit /tag= j
FT FT repeat_unit /note= "Inverted w.r.t. repeat at 3126-3243"
FT FT repeat_unit /tag= k
FT FT repeat_unit /note= "Inverted w.r.t. repeat at 3782-4070"
FT FT repeat_unit /tag= l
FT FT repeat_unit /note= "Inverted w.r.t. repeat 3487-3759"
FT FT repeat_unit /tag= m
FT FT repeat_unit /note= "Inverted w.r.t. repeat 3487-3759"
FT FT repeat_unit /tag= n
FT FT repeat_unit /note= "Inverted w.r.t. repeat 3487-3759"
FT FT repeat_unit /tag= o
FT FT repeat_unit /note= "Inverted w.r.t. repeat 3487-3759"
FT FT repeat_unit /tag= p
FT FT repeat_unit /note= "Inverted w.r.t. repeat 3487-3759"
FT FT repeat_unit /tag= q
FT FT repeat_unit /note= "Inverted w.r.t. repeat 3487-3759"
FT FT repeat_unit /tag= r
FT FT repeat_unit /note= "Inverted w.r.t. repeat 3487-3759"
FT FT repeat_unit /tag= s
FT FT repeat_unit /note= "Inverted w.r.t. repeat 3487-3759"
FT FT repeat_unit /tag= t
FT FT repeat_unit /note= "Inverted w.r.t. repeat 3487-3759"
FT FT repeat_unit /tag= u
FT FT repeat_unit /note= "Inverted w.r.t. repeat 3487-3759"
FT FT repeat_unit /tag= v
FT FT repeat_unit /note= "Inverted w.r.t. repeat 3487-3759"
FT FT repeat_unit /tag= w
FT FT repeat_unit /note= "Inverted w.r.t. repeat 3487-3759"
FT FT repeat_unit /tag= x
FT FT repeat_unit /note= "Inverted w.r.t. repeat 3487-3759"

OS	Caenorhabditis elegans.	Location/Qualifiers
FT	repeat_unit	1356..1472
FT		/*tag= a
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t repeat at 1490-1614"
FT	repeat_unit	1490..1614
FT		/*tag= b
FT		/rpt_type= INVERTED
FT	exon	/note= "Inverted w.r.t. repeat at 1356-1472"
FT		2232..2366
FT		/*tag= c
FT		/number= Exon_1
FT	intron	2367..2429
FT		/*tag= d
FT		/number= Intron_1
FT	exon	2450..2575
FT		/*tag= e
FT		/number= Exon_2
FT		2576..2853
FT		/*tag= f
FT		/number= Intron_2
FT	exon	2854..3107
FT		/*tag= g
FT		/number= Exon_3
FT	intron	3108..4302
FT		/*tag= h
FT		/number= Intron_3
FT	repeat_unit	3126..3243
FT		/*tag= i
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 3329-3396"
FT	repeat_unit	3329..3396
FT		/*tag= j
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 3126-3243"
FT	repeat_unit	3487..3759
FT		/*tag= k
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 3782-4070"
FT	repeat_unit	3782..4070
FT		/*tag= l
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat 3487-3759"
FT	exon	4303..4634
FT		/*tag= m
FT		/number= Exon_4
FT	intron	4635..5546
FT		/*tag= n
FT		/number= Intron_4
FT	repeat_unit	4688..4719
FT		/*tag= o
FT		/rpt_type= INVERTED
FT		/note= "Repeat 3"
FT	repeat_unit	5221..5330
FT		/*tag= p
FT		/rpt_type= INVERTED
FT		/note= "Repeat 3"
FT	exon	5547..5760
FT		/*tag= q
FT		/number= Exon_5
FT	intron	5761..5814
FT		/*tag= r
FT		/number= Intron_5
FT	exon	5815..5942
FT		/*tag= s
FT		/number= Exon_6
FT	intron	5943..6297
FT		/*tag= t
FT		/number= Intron_6
FT	repeat_region	6062..6138
FT		/*tag= u
FT		/rpt_type= INVERTED

Query Match	99.0%; Score 990; DB 9; Length 7653;
Best Local Similarity 99.5%; Pred. No. 0.00e+00;	
Matches 995; Conservative 0; Mismatches 5; Indels 0; Gaps 0;	
Db 1801 ttccaggtgacaaacagaaaacaaacacacacaaacatlttaaaaaatcagtttccaat 1860	FT exon 6298..6537 /*tag= v /number= Exon_7 6535
QY 1801 TTCCAGGCTGACAAACAGAAACAAACACACAAACATTTAAAAATCAGTTTCAAT 1860	FT mutation /*tag= w /note= "G>A, from allele n2426" 6538..7012 /*tag= x /number= Intron_7 6567..6625 /*tag= y /rpt_type= INVERTED /note= "Inverted w.r.t. repeat at 6905-6965" 6905..6965 /*tag= z /rpt_type= INVERTED /note= "Inverted w.r.t. repeat 6567-6625" 7012..7075 /*tag= aa /number= Exon_8
Db 1861 taaaaataacgatttctcatgaaaaatgtgtttatgtttgccaataataaagagaact 1920	FT intron /*tag= x /number= Intron_7 6567..6625 /*tag= y /rpt_type= INVERTED /note= "Inverted w.r.t. repeat 6567-6625" 7012..7075 /*tag= aa /number= Exon_8
QY 1861 TAAAAATAACGATTTCCTCATTTGAAATTTGTGTTTATGTTCGAAAAATAAAAAGAGAAGT 1920	FT repeat_unit /*tag= x /number= Intron_7 6567..6625 /*tag= y /rpt_type= INVERTED /note= "Inverted w.r.t. repeat 6567-6625" 7012..7075 /*tag= aa /number= Exon_8
Db 1921 gattcaaacacattttaacaaaaaaaccccaaatcgccagaaatcaagataaaaaa 1980	FT exon 6298..6537 /*tag= v /number= Exon_7 6535
QY 1921 GATTCAAAACATTTTAACAAAAAAACCCCAAAATTCGCCAGAAATCAAGATAAAAAA 1980	FT mutation /*tag= w /note= "G>A, from allele n2426" 6538..7012 /*tag= x /number= Intron_7 6567..6625 /*tag= y /rpt_type= INVERTED /note= "Inverted w.r.t. repeat 6567-6625" 7012..7075 /*tag= aa /number= Exon_8
Db 1981 ttcaagaggtcacaattttccgatttctactgacttccacttttttcgtagttcagt 2040	FT intron /*tag= x /number= Intron_7 6567..6625 /*tag= y /rpt_type= INVERTED /note= "Inverted w.r.t. repeat 6567-6625" 7012..7075 /*tag= aa /number= Exon_8

Query Match 99.0%; Score 990; DB 9; Length 7653;
 Best Local Similarity 99.5%; Pred. No. 0.00e+00;
 Matches 995; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1801 ttccaggtgacaaacagaaacaaacacaaacattttaaataatcagtttcaaataat 1860
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1801 TTCCAGGCTGACAAACAGAAACAAACACAAACATTTTAAAAATCAGTTTCAAT 1860
 Db 1861 taaaaataacgatttctcatgtgaaattgtgtttatgtttccaaaataaagaagaact 1920
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1861 TAAAAATACGATTTCTCATGTGAAATGTGTTTATGTTCGAAATATAAGAGAACT 1920
 Db 1921 gattcaaaaacaattttaaaaaaaaccccaaatcgcagaatatcaagataaaaaa 1980
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1921 GATCAAAACAATTTTAAACAAAAAAACCCCAAAATTCGCCAGAAATCAAGATAAAAA 1980
 Db 1981 ttcaagaggtcaaaatttccgatttactgacttccaccttttttcgtagtcaagt 2040
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1981 TTCAAGAGGGTCAAAATTTCCGATTTTACTGACTTTCACCTTTTTCGTAGTCAAT 2040
 Db 2041 gcaagttgtgaggtttttgacgaaaaactaggaataaaatcgaataaaattactcaaatcg 2100
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2041 GCAGTGTGTGAGTTTTGTACGAAAACTAGGAAAAAAATCGATAAAAATTTACTCAATCG 2100
 Db 2101 agctgaattttgaggaacaatgtttaaaaaaaacactattttccaataattcactcat 2160
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2101 AGCTGAATTTTGAGGACAAATGTTAAAAAAACACTATTTTCCAATTAATTTCACTCAT 2160
 Db 2161 ttccaactaaatcgaaaaatcaaatcgtaactgactacggtcagtagagagtcacac 2220
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2161 TTTCACTAAATCGAAAAATCAAAATCGTACTCTGACTACGGGTCACTAGAGAGTCAAC 2220
 Db 2221 atcagccgaagatgatgctcaagatagaagagcttgtagagagaacattatgatgt 2280
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2221 ATCAGCCGAGATGATGCGTCAAGATAGAAGAGCTTGTAGAGAGAGACATTAATGATGT 2280
 Db 2281 tctctagtcattcctaaagtgcgtgaataattctcgaagtttctcatcgcaaaacagtggtga 2340
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2281 TCTCTAGTCATCTTAAAGTCGATGAATTTCTCGAAGTTCCTCATCGCAAAACAGTGTGA 2340
 Db 2341 atagtgaataatggagatatgatataatgtgtgagttttaaataataataattttaaataa 2400
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2341 ATAGTGAATATGAGATATGATTAATGTGAGTTTAAATCGAATAATAATTTTAAAAAAA 2400
 Db 2401 aattgataataataaagaataattttgcagtcatagtgaacggttcgcgagaagaagacggg 2460
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2401 AATTGATATATATAAGAAATATTTTGCAGTCATGTGAACGGTTTCGAGAGAGACGGG 2460
 Db 2461 agatcgtgaagcagtgcaacgacgggagatgtggcgttcgacgcgttttatgatgctc 2520
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2461 AGATCGTGAAGCAGTGCACGACGGGAGATGTGGCGTTTCGACGCGTTCATGATGCTC 2520
 Db 2521 ttgcgtctacgggacacgaagagactgtggaagttcttgaaagctctgcagatcgtagg 2580
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2521 TTCGCTCTACGGGACACGAGAGACTTCTGAAGTCTTGAACTCTCGCCAGATCGTAGG 2580
 Db 2581 tttttaagttcggcgcaaaagcgaaggtctcacggaataaagagcgatcgtaatttt 2640
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2581 TTTTAAAGTTCGGCGCAAAAGCAAGGCTCTCACGGAAAAAAGAGCGGATCGTAATTTT 2640
 Db 2641 gcaacccacggcgacggtttttctctcgaaaatacgyaaatatatgcacttcccaatat 2700
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2641 GCAACCCACGGCGACGGTTTTCCTCCGAAAAATCGGAAATTAATGCACTTCCCAAAAT 2700
 Db 2701 ttgaagtgaaatatattattactgaaagctcgagtgatattatttttaacacta 2760
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2701 TTGAAGTGAATATATTTTATTACTGAAAGCTCGAGTGAATTAATTTTAAACACTA 2760
 Db 2761 atttcgtggcgcaaaagccattttgtagatttgccgaa 2800
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 2761 ATTTTCGTGGCGCAAAAGGCCATTTGTAGATTGCGCAA 2800

RESULT 6
 ID 064738 standard; DNA; 7653 BP.
 AC 064738;
 DT 23-JUN-1994 (first entry)
 DE ced-3 (C5940T) gene.
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.
 FH Key
 FT repeat_unit
 FT 1356..1472
 FT location/Qualifiers
 FT /rpt_type= a
 FT /note= "Inverted w.r.t repeat at 1490-1614"
 FT 1490..1614
 FT /rpt_type= b
 FT /note= "Inverted w.r.t. repeat at 1356-1472"
 FT 2232..2366
 FT /rpt_type= c
 FT /number= Exon_1
 FT 2367..2429
 FT /tag= d
 FT /number= Intron_1
 FT 2450..2575
 FT /rpt_type= e
 FT /number= Exon_2
 FT 2576..2853
 FT /tag= f
 FT /number= Intron_2
 FT 2854..3107
 FT /tag= g
 FT /number= Exon_3
 FT 3108..4302
 FT /tag= h
 FT /number= Intron_3
 FT 3126..3243
 FT /tag= i
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3329-3396"
 FT 3329..3396
 FT /tag= j
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3126-3243"
 FT 3487..3759
 FT /tag= k
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3782-4070"
 FT 3782..4070
 FT /tag= l
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 3487-3759"
 FT 4303..4634
 FT /tag= m
 FT /number= Exon_4
 FT 4635..5546
 FT /tag= n
 FT /number= Intron_4
 FT 4688..4719
 FT /tag= o
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT 5221..5330
 FT /tag= p
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT 5547..5760
 FT /tag= q
 FT /number= Exon_5
 FT 5761..5814
 FT /tag= r
 FT /number= Intron_5

Db 2701 ttgaagtgaatatatttactgaagctcgagtgattatttttaacacta 2760
 |||||
 Qy 2701 TTGAAGTGAATATATTTTACTGAAGCTCGAGTGATTTATTTTAACTACTA 2760
 Db 2761 atttcgtggcgcaaaagccattttagatttgcgcaa 2800
 |||||
 Qy 2761 ATTTTCGTGGCGCAAGGCCATTGTAGATTGCCGAA 2800

RESULT 5
 ID 064741 standard; DNA; 7653 BP.
 AC 064741;
 DT 23-JUN-1994 (first entry)
 DE ced-3 (G6372A) gene.
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.
 FH Key
 FT repeat_unit Location/Qualifiers
 FT 1356..1472
 FT /*tag= a
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t repeat at 1490-1614"
 FT 1490..1614
 FT /*tag= b
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1356-1472"
 FT 2232..2366
 FT /*tag= c
 FT /number= Exon_1
 FT 2367..2429
 FT /*tag= d
 FT /number= Intron_1
 FT 2450..2575
 FT /*tag= e
 FT /number= Exon_2
 FT 2576..2853
 FT /*tag= f
 FT /number= Intron_2
 FT 2854..3107
 FT /*tag= g
 FT /number= Exon_3
 FT 3108..4302
 FT /*tag= h
 FT /number= Intron_3
 FT 3126..3243
 FT /*tag= i
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3329-3396"
 FT 3329..3396
 FT /*tag= j
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3126-3243"
 FT 3487..3759
 FT /*tag= k
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3782-4070"
 FT 3782..4070
 FT /*tag= l
 FT /number= Exon_4
 FT 4635..5546
 FT /*tag= m
 FT /number= Intron_4
 FT 4688..4719
 FT /*tag= o
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT 5221..5330
 FT repeat_unit

FT /*tag= p
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT 5547..5760
 FT /*tag= q
 FT /number= Exon_5
 FT 5761..5814
 FT /*tag= r
 FT /number= Intron_5
 FT 5815..5942
 FT /*tag= s
 FT /number= Exon_6
 FT 5943..6297
 FT /*tag= t
 FT /number= Intron_6
 FT 6062..6138
 FT /*tag= u
 FT /rpt_type= INVERTED
 FT 6298..6537
 FT /*tag= v
 FT /number= Exon_7
 FT 6372
 FT /*tag= w
 FT /note= "G>A, from allele n1286"
 FT 6538..7012
 FT /*tag= x
 FT /number= Intron_7
 FT 6567..6625
 FT /*tag= y
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 6905-6965"
 FT 6905..6965
 FT /*tag= z
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 6567-6625"
 FT 7012..7075
 FT /*tag= aa
 FT /number= Exon_8
 FT 709325685-A.
 FT 23-DEC-1993.
 FT 14-JUN-1993; U05701.
 FT 12-JUN-1992; US-897788.
 FT 20-NOV-1992; US-979638.
 FT (MASI) MASSACHUSETTS INST TECHNOLOGY.
 FT Horvitz HR, Shaham S, Yuan J;
 FT WPI; 94-007542/01.
 FT P-PSDB; R53284.
 FT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 FT develop agents to increase or prevent cell death in organisms
 FT Claim 14; Fig 4; 127pp; English.
 FT The sequences given in Q64735-45 represent mutations of the C. elegans
 FT ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
 FT was most abundant in embryos, but was also detected in larvae and young
 FT adults, suggesting that ced-3 is expressed not only in cells undergoing
 FT cell death. The four largest introns as well as sequences 5' of the
 FT start codon contain repetitive elements, some of which have been
 FT characterised in non-coding regions of other C. elegans genes, such
 FT as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
 FT length. Ced-3 is highly hydrophilic with no significant hydrophobic
 FT region that might be a transmembrane region. One region of Ced-3 is
 FT very rich in serine. It is thought that this region is involved in
 FT protein-protein interactions, similar to acid blobs in transcription
 FT factors. Of the mutations which occur within the ced-3 gene, eight of
 FT the mutations are missense mutations, two are nonsense mutations and
 FT two are putative splicing mutations. These mutations establish the
 FT null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
 FT function is not essential for viability. The ced-3 and ced-4 gene
 FT products may be used to develop agents for treating conditions
 FT characterised by cell deaths, such as myocardial infarction, stroke,
 FT degenerative disease, traumatic brain injury, hypoxia, pathogenic
 FT infection, aging or hair loss.
 FT Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

FT		/number= Exon_4
FT	intron	4635..5546
FT		/*tag= n
FT		/number= Intron_4
FT	repeat_unit	4688..4719
FT		/*tag= o
FT		/rpt_type= INVERTED
FT		/note= "Repeat 3"
FT	repeat_unit	5221..5330
FT		/*tag= p
FT		/rpt_type= INVERTED
FT		/note= "Repeat 3"
FT	exon	5547..5760
FT		/*tag= q
FT		/number= Exon_5
FT	intron	5761..5814
FT		/*tag= r
FT		/number= Intron_5
FT	exon	5815..5942
FT		/*tag= s
FT		/number= Exon_6
FT	intron	5943..6297
FT		/*tag= t
FT		/number= Intron_6
FT	repeat_region	6062..6138
FT		/*tag= u
FT		/rpt_type= INVERTED
FT	exon	6298..6537
FT		/*tag= v
FT		/number= Exon_7
FT	mutation	6434
FT		/*tag= w
FT		/note= "C>T, from allele n1129 and n1164"
FT	intron	6538..7012
FT		/*tag= x
FT		/number= Intron_7
FT	repeat_unit	6567..6625
FT		/*tag= y
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 6905-6965"
FT	repeat_unit	6905..6965
FT		/*tag= z
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat 6567-6625"
FT	exon	7012..7075
FT		/*tag= aa
FT		/number= Exon_8
PN	WO9325685-A.	
PD	23-DEC-1993.	
PF	14-JUN-1993; U05701.	
PR	12-JUN-1992; US-897788.	
PR	20-NOV-1992; US-979638.	
PA	(MASI) MASSACHUSETTS INST TECHNOLOGY.	
PI	Horvitz HR, Shaham S, Yuan J;	
DR	WPI; 94-007542/01.	
DR	P-PSDB; R53285.	
PT	Isolated C elegans cell death genes ced-3 and ced-4 - used to	
PT	developed agents to increase or prevent cell death in organisms	
PS	Claim 14; Fig 4; 127pp; English.	
CC	The sequences given in Q64735-45 represent mutations of the C. elegans	
CC	ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and	
CC	was most abundant in embryos, but was also detected in larvae and young	
CC	adults, suggesting that ced-3 is expressed not only in cells undergoing	
CC	cell death. The four largest introns as well as sequences 5' of the	
CC	start codon contain repetitive elements, some of which have been	
CC	characterised in non-coding regions of other C. elegans genes, such	
CC	as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in	
CC	length. Ced-3 is highly hydrophilic with no significant hydrophobic	
CC	region that might be a transmembrane region. One region of Ced-3 is	
CC	very rich in serine. It is thought that this region is involved in	
CC	protein-protein interactions, similar to acid blobs in transcription	
CC	factors. Of the mutations which occur within the ced-3 gene, eight of	
CC	the mutations are missense mutations, two are nonsense mutations and	

CC	two are putative splicing mutations. These mutations establish the
CC	null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC	function is not essential for viability. The ced-3 and ced-4 gene
CC	products may be used to develop agents for treating conditions
CC	characterised by cell deaths, such as myocardial infarction, stroke,
CC	degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC	infection, aging or hair loss.
CC	Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;
QY	Query Match 99.0%; Score 990; DB 9; Length 7653;
QY	Best Local Similarity 99.5%; Pred. No. 0.00e+00;
QY	Matches 995; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db	1801 ttccaggtcgtacaacagaaacaaacacacaacaacattttaaaaaatcagtttccaat 1860
QY	1801 TTCCAGGCTGACAAACAGAAACAAACACAAACATTTAAAAATCAGTTTCAAAT 1860
Db	1861 taaaaaaacgatttctcatgtgaaaaattgtgtttatgtttgcacaaataaagaagaact 1920
QY	1861 TAAAAATTAACGATTTCATTTGAAATTTGTTTATGTTTGGCAAAATAAAAAGAGAACT 1920
Db	1921 gatcaaaaacaattttaacaaaaaaaccccaaaatcgcagaataatcaagataaaaaa 1980
QY	1921 GATTCAAAACAATTTTAACAAAAAAAACCCCAAAATTCGCAGAAATCAAGATAAAAAA 1980
Db	1981 ttcaagaggtcaaaatttccgattttactgactttcaccttttttcgtagttcagt 2040
QY	1981 TTCAAGAGGGTCAAAATTTCCGATTTTACTGACTTTCACCTTTTTCGTAGTTCAGT 2040
Db	2041 gcaagtgttgaggttttttgacgaaaaactaggaaaaaaatcgataaaaattactcaatcg 2100
QY	2041 GCAGTTGTTGGAGTTTTCGACGAAACTAGGAAAAAATCGATAAAAATTACTCAAAATCG 2100
Db	2101 agctgaaattttgaggaacaatgttttaaaaaaaacactatlttccaataattcactcat 2160
QY	2101 AGCTGAATTTTGAGGACCAATGTTAAAAAAACACTATTTTCCATAATTTCACTCAT 2160
Db	2161 ttccaagactaaatcgaaaatcaaatcgtactctgactacggtcagtagagaagtcacac 2220
QY	2161 TTTCAGACTAAATCGAAAATCAAAATCTGACTCTGACTACGGGTAGAGAGGTCACACC 2220
Db	2221 atcagccgaagatgatgctgcgtcagaatagaagagcttctagagaaggaacattatgagt 2280
QY	2221 ATCAGCCGAAGATGATGCGTCAAGTAGAAGGAGCTTGCTAGAGAGAACATTATGATGT 2280
Db	2281 tctcgaatcatctaaagaatcgatgaaattctcgaagtttctcatcgcaaaaaaagtggtga 2340
QY	2281 TCTCTAGTCACTCTAAAAGTCGATGAATTTCTCGAAGTTCTCATCGCAAAACAAGTGTGA 2340
Db	2341 atagtataatgtagatatgattaaatgtgagttttaatcgaataaattttaaaaaaa 2400
QY	2341 ATAGTGAATAATGAGATATGATTAATGTGAGTTTAAATCGAATAATAATTTAAAAAAA 2400
Db	2401 aattgataataataaagaatatltttgcagtcattgtgaaacggttcgcgagaagagacggg 2460
QY	2401 AATTGATAATAATAAAGATATTTTTCAGTCATGTGGAACGGTTCGCAGAGAAGACGGG 2460
Db	2461 agatcgtgaaaaagcagtcgcaacgacggygagatgtgycgtttccaacgcygtttatgatgtc 2520
QY	2461 AGATCGTGAAGAAGCAGTGCACGACGGGAGATGTGGCGTTCGACGCGTTTTATGATGCTC 2520
Db	2521 ttccgtctacggaacacgaagactgtctgaagttcttgaagctctcgccagatcgttagg 2580
QY	2521 TTCCGCTACGGGACACGAAGGACTTGCTGAAGTCTTGAACTCTTCGCCAGATCGTAGG 2580
Db	2581 tttttaaaagttcgcgcgcaaaagcaagggtctcacgaaaaaaagagccgataat 2640
QY	2581 TTTTAAAGTTTCGGCGCAAAAGCAAGGGTCTCACGGAATAAAGAGCGGATCGTAATTTT 2640
Db	2641 gcaaccacccggcagcgttttttctctcggaaaaatcggaatatatgacatttcccaatat 2700
QY	2641 GCAACCCACCGGCACGGTTTTTCCTCCGAAATAATCGAATAATATGACACTTCCCAATAT 2700

CC contain repetitive elements, some of which have been characterised
CC in non-coding regions of other C. elegans genes, such as fem-1, lin-12
CC and myoD. The Ced-3 protein is 503 amino acids in length. Ced-3 is
CC highly hydrophilic with no significant hydrophobic region that might
CC be a transmembrane region. One region of Ced-3 is very rich in serine.
CC It is thought that this region is involved in protein-protein inter-
CC actions, similar to acid blobs in transcription factors. Of the
CC mutations which occur within the ced-3 gene, eight of the mutations
CC are missense mutations, two are nonsense mutations and two are putative
CC splicing mutations (see also Q64735-45). These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2429 A; 1453 C; 1272 G; 2499 T;

Query Match 99.4%; Score 994; DB 9; Length 7653;
Best Local Similarity 99.7%; Pred. No. 0.00e+00;
Matches 997; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 1801 ttccaggtcgaacaaacagaacaaacacacacacatttaaaatcagtttcaaat 1860
|||
QY 1801 TTCAGGCTGACAAACAGAAACAAACACAAACATTTAAATAATCAGTTTCAAT 1860
1861 taaaataacagatttctcatgtgaatgtgtttatgtttgcgaataaagaagaact 1920
|||
QY 1861 TAAAAATAACGATTTCTCATTTGAAATTTGTGTTTATGTTGCCAAATAAAGAGACT 1920
Db 1921 gatcaaacacatttaacacacacacacacacacattccgcaagaataatcaagataaaaa 1980
|||
QY 1921 GATCAAAACAAATTTTACCAAAAAAACCCCAAAATTCGCCAGAAATCAAGATAAAAA 1980
Db 1981 ttcaagaggtcacaattttccgatttactgtacttccacttttttccgtagttcagt 2040
|||
QY 1981 TTCAAGAGGCTCAAAATTTCCGATTTTACTGACTTTCACCTTTTTCGTAGTTCACT 2040
Db 2041 gcaattgttggagtttttgacgaaaactaggaacaaatcgataaaatctactcaaatcg 2100
|||
QY 2041 GCAGTGTGTGAGTTTTGTACGAAACTAGGAAAAAATCGATAAAAAATTACTCAATCG 2100
Db 2101 agctgaattttgaggaacatgtttaaaaaaacacatattttccaataattccactcat 2160
|||
QY 2101 AGCTGAATTTTGAGGACAATGTTTAAAAAACACACTATTTTCCAATAATTTCACTCAT 2160
Db 2161 ttccagactaaatcgaaatcaaatcgtactctgactaggtcagtagagagtcacac 2220
|||
QY 2161 TTTCAGACTAAATCGAAATCAATCGTACTCTGACTACGGGTCACTAGAGAGGTCAACC 2220
Db 2221 atcagccgaagatgtagtgcacgaagatagagagctgtctagagaggaacattatgtat 2280
|||
QY 2221 ATCAGCCGAAGATGATGCGTCAAGATAGAAGAGGCTTGTCTAGAGGAACATTAATGATGT 2280
Db 2281 tctctagtcattctaaagtgcataaattctcgaagttctcattcgcaaaacagttgtga 2340
|||
QY 2281 TCTCTAGTCATCTAAAGTCCATGAATTTCTCGAAGTTCTCATCGCAAAACAGTGTGA 2340
Db 2341 atagtataatgagagatatgaatgtgagtttttaatacgataataatttaaaaaa 2400
|||
QY 2341 ATAGTGATTAATGAGATATGATTAATGTGAGTTTAAATCGAATAATTAATTAATAAAAA 2400
Db 2401 aattgataataataagaataatttgcagtcattggaacggttcgcgagaagagcggg 2460
|||
QY 2401 AATTGATTAATATAAAGAAATTTTTCAGTCAATGTGAACGGTTCGCGAGAGAGACGGG 2460
Db 2461 agatcgtgaaagcagtgcaacgacgagggagatgtggttcgcgcgttttatgtatgctc 2520
|||
QY 2461 AGATCGTGAAGCAGTGCACGACGGGAGATGTGGGCTTCGACGCGTTTATGATGCTC 2520
Db 2521 ttgcctctacgagacgagagactgtgtaagtctctgaacctctgcacagatcgttagg 2580
|||
QY 2521 TTGCTCTACGAGACAGAGGACTTGTGAAGTCTTGAACCTCTCGCCAGATCGTAGG 2580

Db 2581 tttaaagtctgcgcgaacaaagcaaggtctccacggaacaaagagccgcatcgtaat 2640
|||
QY 2581 TTTTAAAGTTCGGCGCAAAAGCAAGGGTCTACGGAATAAGAGCGGATCGTAATTT 2640
Db 2641 gcaaccaccgcaaggtttttctcctccgaaatcggaattatgcactttcccaatat 2700
|||
QY 2641 GCAACCCACCGCAGCGTTTTCTCCCGAAATCGGAATTAATGCACTTCCCAATAT 2700
Db 2701 ttgaagtgaatatatttattactgaaagctcagtgatattatatttttaacacta 2760
|||
QY 2701 TTGAAGTGAATATATTTTATTACTGAAGCTCGAGTGAATTAATTTTAAACACTA 2760
Db 2761 atttcgtgcgcaaaagccattttagatttgcgaa 2800
|||
QY 2761 ATTTCTGTGGCGCAAAAGCCATTTGTAGATTGCGCGAA 2800

RESULT 4
ID Q64742 standard; DNA; 7653 BP.
AC Q64742;
DT 23-JUN-1994 (first entry)
DE ced-3 (C6434T) gene.
KW C. elegans; ced-4; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
FH Caenorhabditis elegans.
FH Key
FH Location/Qualifiers
FT repeat_unit
FT 1356..1472
FT /*tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT 1490..1614
FT /*tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT 2232..2366
FT /*tag= c
FT /number= Exon_1
FT 2367..2429
FT /*tag= d
FT /number= Intron_1
FT 2450..2575
FT /*tag= e
FT /number= Exon_2
FT 2576..2853
FT /*tag= f
FT /number= Intron_2
FT 2854..3107
FT /*tag= g
FT /number= Exon_3
FT 3108..4302
FT /*tag= h
FT /number= Intron_3
FT 3126..3243
FT /*tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT 3329..3396
FT /*tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT 3487..3759
FT /*tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT 3782..4070
FT /*tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT 4303..4634
FT /*tag= m

DT 23-JUN-1994 (first entry)
 DE ced-3 gene.
 KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
 KW embryogenesis; cell death; hydrophilic; transmembrane; region;
 KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
 KW protein synthesis; null phenotype; calcium-binding domain; ss.
 OS Caenorhabditis elegans.
 FH key Location/Qualifiers
 FT repeat_unit 1356..1472
 FT /tag= a
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1490-1614"
 FT 1490..1614
 FT /tag= b
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 1356-1472"
 FT 2232..2366
 FT /tag= c
 FT /number= Exon_1
 FT 2310
 FT /tag= d
 FT /note= "C>T, from allele n1040"
 FT 2367..2429
 FT /tag= e
 FT /number= Intron_1
 FT 2430..2575
 FT /tag= f
 FT /number= Exon_2
 FT 2487
 FT /tag= g
 FT /note= "G>A, from allele n718"
 FT 2576..2853
 FT /tag= h
 FT /number= Intron_2
 FT 2854..3107
 FT /tag= i
 FT /number= Exon_3
 FT 3108..4302
 FT /tag= j
 FT /number= Intron_3
 FT 3126..3243
 FT /tag= k
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3329-3396"
 FT 3329..3396
 FT /tag= l
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3126-3243"
 FT 3487..3759
 FT /tag= m
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 3782-4070"
 FT 3782..4070
 FT /tag= n
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 3487-3759"
 FT 4303..4634
 FT /tag= o
 FT /number= Exon_4
 FT 4635..5546
 FT /tag= p
 FT /number= Intron_4
 FT 4688..4719
 FT /tag= q
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT 5221..5330
 FT /tag= r
 FT /rpt_type= INVERTED
 FT /note= "Repeat 3"
 FT 5547..5760
 FT /tag= s
 FT /number= Exon_5

FT mutation 5757
 FT /tag= t
 FT /note= "G>A, from allele n2433"
 FT 5761..5814
 FT /tag= u
 FT /number= Intron_5
 FT 5815..5942
 FT /tag= v
 FT /number= Exon_6
 FT 5940
 FT /tag= w
 FT /note= "C>T, from allele n1165"
 FT 5943..6297
 FT /tag= x
 FT /number= Intron_6
 FT 6062..6138
 FT /tag= y
 FT /rpt_type= INVERTED
 FT 6297
 FT /tag= z
 FT /note= "G>A, from allele n717"
 FT 6298..6537
 FT /tag= aa
 FT /number= Exon_7
 FT 6322
 FT /tag= ab
 FT /note= "C>T, from allele n1949"
 FT 6372
 FT /tag= ac
 FT /note= "G>A, from allele n1286"
 FT 6434
 FT /tag= ad
 FT /note= "C>T, from alleles n1129 and n1164"
 FT 6485
 FT /tag= ae
 FT /note= "C>T, from allele n2430"
 FT 6535
 FT /tag= af
 FT /note= "G>A, from allele n2426"
 FT 6538..7012
 FT /tag= ag
 FT /number= Intron_7
 FT 6567..6625
 FT /tag= ah
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat at 6905-6965"
 FT 6905..6965
 FT /tag= ai
 FT /rpt_type= INVERTED
 FT /note= "Inverted w.r.t. repeat 6567-6625"
 FT 7013..7075
 FT /tag= aj
 FT /number= Exon_8
 FT 7020
 FT /tag= ak
 FT /note= "C>T, from allele n1163"
 FT mutation
 FT WO9325685-A.
 FT 23-DEC-1993.
 FT 14-JUN-1993.
 FT 12-JUN-1992; US-897788.
 FT 20-NOV-1992; US-979638.
 FT (MASI) MASSACHUSETTS INST TECHNOLOGY.
 FT Horvitz HR, Shahan S, Yuan J;
 FT WPI; 94-007542/01.
 FT P-PSDB; R47466.
 FT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 FT develop agents to increase or prevent cell death in organisms
 FT Claim 2; Fig 4; 127P; English.
 FT This sequence represents the C. elegans ced-3 gene. A 2.8 kb mRNA
 FT was identified as the ced-3 transcript and was most abundant in
 FT embryos, but was also detected in larvae and young adults, suggesting
 FT that ced-3 is expressed not only in cells undergoing cell death. The
 FT four largest introns as well as sequences 5' of the start codon


```
FT intron 5943..6297
FT /tag= x
FT /label= Intron-6
FT repeat_unit 6062..6138
FT /tag= y
FT /label= Repeat-4
FT exon 6298..6537
FT /tag= z
FT allele 6322
FT /tag= aa
FT /label= T(n1949)
FT /note= "creates premature stop codon at 412"
FT allele 6372
FT /tag= ab
FT /label= A(n1286)
FT allele 6434
FT /tag= ac
FT /label= T(n1129,n1164)
FT /note= "causes A449V mutation"
FT allele 6485
FT /tag= ad
FT /label= T(n2430)
FT /note= "causes A466V mutation"
FT allele 6535
FT /tag= ae
FT /label= A(n2426)
FT /note= "causes E483K mutation"
FT intron 6538..7012
FT /tag= af
FT /label= Intron-7
FT repeat_unit 6567..6625
FT /tag= ag
FT /label= Repeat-5
FT repeat_unit 6905..6965
FT /tag= ah
FT /label= Repeat-5
FT exon 7013..7653
FT /tag= ai
FT /note= "in-frame stop codon at 7073-7075"
FT allele 7020
FT /tag= aj
FT /label= T(n1163)
FT /note= "causes S486F mutation"
FT W09625946-A1.
FT 23-AUG-1996.
FT 23-FEB-1996; U02473.
FT 24-FEB-1995; US-394189.
FT (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT Horvitz HR, Shaham S, Yuan J;
FT WPI: 96-425082/42.
FT P-PSDB; R98754.
FT Ced-3 and human interleukin 1-beta convertase genes and proteins -
FT useful to treat inflammation and diseases characterised by cell
FT death
FT Claim 18; Fig 3; 139pp; English.
FT The Caenorhabditis elegans ced-3 gene (T38196) was cloned by
FT mapping RFLPs and chromosome walking, and genomic DNA cloned
FT in plasmid pJ107 was sequenced. EMS-induced alleles were also
FT sequenced. The gene codes for a cell death protein (R98754) that
FT is structurally similar to human interleukin-1 beta converting
FT enzyme (ICE) (R98755), suggesting that Ced-3 protein may be a
FT cysteine protease like ICE and that ICE may be a human equivalent
FT of the nematode cell death gene. The ced-3 gene can be used as a
FT probe or in the prodn. of Ced-3 protein and novel drugs for
FT enhancing or inhibiting the activity of ICE, ced-3 and related
FT genes for the treatment of inflammatory diseases and/or diseases
FT caused by cell death. Novel inhibitors of ced-3 activity include
FT portions of the ced-3 gene and its product.
FT Sequence 7653 BP; 2429 A; 1456 C; 1270 G; 2498 T;
```

Query Match 100.0%; Score 1000; DB 24; Length 7653;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db 1801 ttccaggtgacaaacagaacaacaacaacacattttaaaaaatcagtttcaaat 1860
QY 1801 TTCAGGCTGACAAACAGAAACAAAAACAAACAAACATTTTAAAAATCAGTTTCAAAAT 1860
Db 1861 taaaataacgatttctcatlgaanaattgtgtttatgtttgcgaaaataaagagaact 1920
QY 1861 TAAAAATACGATTTCTCATTTGAAAAATGTGTTTATGTTGCCAAAAATAAAGAGAACT 1920
Db 1921 gattcaaaaacaattttaacaaaaaaacccccaaaatttcgccagaataatcaagataaaaa 1980
QY 1921 GATTCAAAACATTTTAAACAAAAAAACCCCAAAATTCGCCAGAAATCAAGATAAAAAA 1980
Db 1981 ttcaagaaggtcaaaaatttccgatttactgacttccaccttttttctgtagtcagt 2040
QY 1981 TTCAAGAGGGTCAAAAATTTCCGATTTTACTGACTTTCACCTTTTTCGTAGTTCAGT 2040
Db 2041 gcagttgtgagatttttgacgaaaactaggaaaaaaatcgataaaaaattactcaaatcg 2100
QY 2041 GCAGTTGTTGAGTTTGTGACGAAAAACTAGGAAAAAAATCGATAAAATTACTCAAAATCG 2100
Db 2101 agctgaattttgagagacaatgttttaaaaaaaacactattttccaataattcactcat 2160
QY 2101 AGCTGAATTTTGAGACAAATGTTTAAAAAAAACACTATTTTCCAAATAATTTCACCTCAT 2160
Db 2161 ttccagactaaatcgaanaatcacaatcgtactctgactacggtcagtagagaggtcaacc 2220
QY 2161 TTCCAGACTAAATCGAAATCAAAATCGTACTCTGACTACGGGTCACTAGAGAGTCAACC 2220
Db 2221 atcagccgaagatgatgctgcgaagatagaaggagcttgcctagagaaggaaacattatgatgt 2280
QY 2221 ATCAGCCGAAGATGATGCGTCAAGATAGAGAGAGCTTGCTAGAGAGAACATTATGATGT 2280
Db 2281 tctctagtcactaaagtcgatgaattctcgaagttctcatcgcaaaaacagtggtga 2340
QY 2281 TCTTAGTCATCTAAAGTCGATGAATTTCTGAAAGTTCTCATCGCAAAAACAAAGTGTGA 2340
Db 2341 atagtataatgagatatgatataatgtgagttttaatcgaataataatttaaaaaa 2400
QY 2341 ATAGTGATAATGAGATATGATTAATGTGAGTTTAAATCGAATAATTTTAAAAAAA 2400
Db 2401 aattgataataaagaatatattttgcagtcagtcatgtgaaacggttcgcgagaagagacggg 2460
QY 2401 AATTGATATATATAAGATATTTTTCAGTCACTGTGGAACGGTTCGCGAAGAAGACGGG 2460
Db 2461 agatcgtgaagcagtcgcaacgacggtgagatgtggttcgacggttttatgatgtc 2520
QY 2461 AGATCGTGAAAGCAGTCGAACGACGGGGAGATGTGGCGTTCGACGCGTTTATGATGCTC 2520
Db 2521 ttgcgtctacgggacacgaaggactgtcgaagttcttgaacctctgcagatcgtag 2580
QY 2521 TTCGCTCTACGGGACACGAAGACTTGTGAAGTTCTTGAACCTCTGCCAGATCGTAGG 2580
Db 2581 tttttaaagttcggcgcaaaagcaagaggttcacggaaaaaaagagcggtatattt 2640
QY 2581 TTTTAAAGTTTCGGCGCAAAAGCAAGGTCCTCACGGAATAAAGAGCGGATCGTAATTT 2640
Db 2641 gcaaccacacgacggtttttcctcgaanaatcggaataatgacattcccaaatat 2700
QY 2641 GCAACCCACCGGACGGTTTTCTCTCCGAAAAATCGGAATAATATGCACTTCCCAATAT 2700
Db 2701 ttgaagtgaatatatttattactgaagctcagtgatattattttttaacacta 2760
QY 2701 TTGAAGTGAATATATTTTATTACTGAAGCTCGAGTGATTTATTTTAAACACTA 2760
Db 2761 atttcgtgcgcaaaagccatttgttagatttcgcaa 2800
QY 2761 ATTTTCGTGGCGCAAAAGGCCATTGTGTAGATTGCGCGAA 2800
```

RESULT 3
ID Q54666 standard; DNA; 7653 BP.
AC Q54666;

|||||
QY 1801 TTCCAGGCTGACAAACAGAAACAAAACACAAACAAATTTAAAAATCAGTTTCAAT 1860
Db 1861 taaaaataacgaattctcatgaaaattgtgtttatgtttcgaaaataaaagagaact 1920
QY 1861 TAAAAATTAACGATTTCTCATTTGAATAATGTGTTTATGTTCGAAAATAAAAAGAGACT 1920
Db 1921 gattcaaaacaattttaacaaaaaaacccccaaaattcgccagaatacaagataaaaaa 1980
QY 1921 GATCAAAACAAATTTTAACAAAAAAACCCTCAAAATTCGCCAGAAATCAAGATAAAAA 1980
Db 1981 ttcaagagggccaattttccgattttactgactttcaccttttttcgtagtcaagt 2040
QY 1981 TTCAAGAGGGTCAAAATTTCCGATTTTACTGACTTTCACCTTTTTCGTAGTTCAGT 2040
Db 2041 gcagttgttgaggtttttgacgaaaaactaggaaaaaaatcgataaaaaattactcaatcg 2100
QY 2041 GCAGTTGTTGAGGTTTTCGACGAAAACTAGGAAAAAAATCGATAAAATTTACTCAATCG 2100
Db 2101 agctgaattttgagacaatgtttaaaaaaaacactattttccaataattcactcat 2160
QY 2101 AGCTGAATTTTGAGGACAAATGTTAAAAAAAACACTATTTTCCAAATTAATTTCACTCAT 2160
Db 2161 ttccagactaaatcgaaaaatcaaatcgtaactctgactacggtcagtagagagtcaccc 2220
QY 2161 TTTCAAGACTAAATCGAAAAATCAAAATCGTACTCTGACTACGGGTCAAGTAGAGAGTCAACC 2220
Db 2221 atcagccgaagatgtatgctcaagatagaagagcttgcttagagaggaacattatgagt 2280
QY 2221 ATCAGCCGAAGATGATGCGCTCAAGATAGAAGGAGCTTGCTAGAGAGGAACATTATGATGT 2280
Db 2281 tctctagtcattctaaaaagtcgaataatctcgaagttctcatcgcaaaaacaaagtgtga 2340
QY 2281 TCTCTAGTCACTTAAAGTCTGAATTAATCTCGAAGTTCTCATCGCAAAACAAGTTTGA 2340
Db 2341 atagtataatggaatatatgataatgtgagttttaatcgaaataaattttaaaaaaa 2400
QY 2341 ATAGTGAATATGAGATATGATTAATGTGAGTTTAAATCGAATAATTAATTTTAAAAAAA 2400
Db 2401 aattgataataaaagaatatatttcgagtcagtcatgtggaacggttcgagagaagacggg 2460
QY 2401 AATTGATTAATATAAAGAATATTTTGCAGTCATGTGAACGTTCCGCGAGAAGACGGG 2460
Db 2461 agatcgtgaaagcagtcgaacgaggggagatgtggtcgttcgaacggttttatgagctc 2520
QY 2461 AGATCGTGAAGCAGTGCACACGACGGGAGATGTGGCGTTCGACGGCTTTATGATGCTC 2520
Db 2521 ttgcctctacgaggaacgaaggaacttgctgaagttcttgaacctctcgccagatcgtagg 2580
QY 2521 TTCGCTCTACGGGACACGAAGGACTTGCTGAAGTCTTGAACCTCTCGCCAGATCGTAGG 2580
Db 2581 tttttaaaagttcggcgcaaaaggggtctcaacgaaaaaaagggcgatcgtaatttt 2640
QY 2581 TTTTAAAGTTTCGGCGCAAAAGGGTCTCACGGAATAAGAGGCGGATCGTAATTTT 2640
Db 2641 gcaaccacccgagcaggtttttctccgaaaatcgaaattatgcactttcccaatat 2700
QY 2641 GCAACCCACCGGACCGGTTTTCCTCCGAAATCGGAATTAAGCACTTCCCAATAT 2700
Db 2701 ttgaagtgaatatattttacttgaaagctcgagtgattatttttttaacacta 2760
QY 2701 TTGAAGTGAATATATTTTATTTACTGAAAGCTCGAGTGATTAATTTTAAACACTA 2760
Db 2761 atttcgtggcgcaaaagggcattttgtagatttgccgaa 2800
QY 2761 ATTTCGTGGCGCAAAAGGCCATTTGTAGATTGCCGAA 2800

RESULT 2
ID T38196 standard; DNA; 7653 BP.
AC T38196;
DT 17-DEC-1996 (first entry)
DE Nematode Ced-3 gene.

KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;
KW cell death; apoptosis; neural degeneration; inflammation;
KW antiinflammatory; ds.
OS Caenorhabditis elegans.
FH Key location/Qualifiers
FT repeat_unit 1356..1472
FT FT
FT /tag= a
FT /label= Repeat-1
FT 1490..1614
FT FT
FT /tag= b
FT /label= Repeat-2
FT 2167..2366
FT FT
FT /tag= c
FT /codon_start= 2232..2234
FT 2310
FT FT
FT /tag= d
FT /label= T(n1040)
FT /note= "causes U27F mutation"
FT 2367..2429
FT FT
FT /tag= e
FT /label= Intron-1
FT 2430..2575
FT FT
FT /tag= f
FT 2487
FT FT
FT /tag= g
FT /label= A(n1718)
FT /note= "causes G65R mutation"
FT 2576..2853
FT FT
FT /tag= h
FT /label= Intron-2
FT 2854..3107
FT FT
FT /tag= i
FT 3108..4302
FT FT
FT /tag= j
FT /label= Intron-3
FT 3126..3243
FT FT
FT /tag= k
FT /label= Repeat-1
FT 3329..3396
FT FT
FT /tag= l
FT /label= Repeat-1
FT 3487..3759
FT FT
FT /tag= m
FT /label= Repeat-2
FT 3782..4070
FT FT
FT /tag= n
FT /label= Repeat-2
FT 4303..4633
FT FT
FT /tag= o
FT 4634..5546
FT FT
FT /tag= p
FT /label= Intron-4
FT 4688..4719
FT FT
FT /tag= q
FT /label= Repeat-3
FT 5221..5330
FT FT
FT /tag= r
FT /label= Repeat-3
FT 5546..5760
FT FT
FT /tag= s
FT 5757
FT FT
FT /tag= t
FT /label= A(n2433)
FT /note= "causes G360S mutation"
FT 5761..5814
FT FT
FT /tag= u
FT /label= Intron-5
FT 5815..5942
FT FT
FT /tag= v
FT 5940
FT FT
FT /tag= w
FT /label= T(n1165)
FT /note= "creates premature stop codon at 403"

[illegible]

FT	/note= "G>A, from allele n1286"
FT	6434
FT	/tag= ad
FT	/note= "C>T, fromm alleles n1129 and n1164"
FT	6485
FT	/tag= ae
FT	/note= "C>T, from allele n2430"
FT	6535
FT	/tag= af
FT	/note= "G>A, from allele n2426"
FT	6538..7012
FT	/tag= ag
FT	/number= Intron_7
FT	6567..6625
FT	/tag= ah
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 6905-6965"
FT	6905..6965
FT	/tag= ai
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat 6567-6625"
FT	7013..7075
FT	/tag= aj
FT	/number= Exon_8
FT	7020
FT	/tag= ak
FT	/note= "C>T, from allele n1163"
FT	1..5850
FT	/tag= al
FT	/note= "Claim 2, inhibitor fragment"
FT	1..3020
FT	/tag= am
FT	/note= "Claim 2, inhibiotr fragment"
PN	WO9325694-A.
PD	23-DEC-1993.
PF	14-JUN-1993; U05705.
PR	12-JUN-1992; US-897788.
PA	20-NOV-1992; US-984182.
PR	(MASI) MASSACHUSETTS INST TECHNOLOGY.
PI	Horvitz HR, Shaham S, Yuan J;
DR	WPI; 94-007551/01.
DR	P-PSDB: R45262.
PT	Agents which affect activity of cell death genes - used to
PT	develop drugs for treating conditions characterised by cell death
PT	or proliferation
PS	Disclosure; Fig 3: 132pp; English.
CC	This sequence encodes the C.elegans cell death gene, ced-3. Fragments
CC	of the amino terminal of the protein encoded by this sequence act as
CC	inhibitors of ced-3. This gene has considerable similarity to human
CC	interleukin-1beta convertase (ICE), which converts pro-interleukin-
CC	1beta to the active cytokine and is involved in inflammatory response
CC	in humans. The similarity between the two sequences suggests that
CC	inhibitors of ced-3 may also act as inhibitors of ICE. Human ICE and
CC	nematode Ced-3 proteins have an overall amino acid similarity of 28%.
CC	The ced-3 inhibitors may be used for identifying agents which affect
CC	the activity of a gene belonging to the to the ced-3/ICE family of
CC	genes and for diagnosis of diseases characterised by cell death. They
CC	can also be used to develop drugs for treating conditions characterised
CC	by cell deaths such as myocardial infarction, stroke, degenerative
CC	disease, traumatic brain injury, hypoxia, pathogenic infection, or
CC	hair loss, or drugs for reducing the proliferative capacity or size
CC	of a population of cells such as cancerous cells, cells which produce
CC	autoreactive antibodies, infected cells, hair follicle cells or cells
CC	which are critical to the life of a parasite, pest or recombinant
CC	organism. They may also be used in the diagnosis of inflammatory
CC	disease.
SO	Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match	100.0%;	Score 1000;	DB 9;	Length 7653;
Best Local Similarity	100.0%;	Pred. No. 0.00e+00;		
Matches 1000;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

(TM)

MPsrch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Fri Aug 6 04:11:21 1999; Maspar time 234.72 Seconds
912.897 Million cell updates/sec

Tabular output not generated.

Title: >US-08-287-669-18
Description: (1801-2800) from US08287669.seq (3 of 10)
Perfect Score: 1000
N.A. Sequence: 1801 TTCACGGCTGACAAACAGAA.....CATTTGTAGATTGCCGAA 2800
Comp: AAGTCGCGACTGTTGTCTT.....GTAACAATCTAACGCGCTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 9.004; Variance 6.867; scale 1.311

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description	Pred. No.
1	1000	100.0	7653	9	Q54401	ced-3 gene.
2	1000	100.0	7653	24	T38196	Nematode Ced-3 gene.
3	994	99.4	7653	9	Q54666	ced-3 gene.
4	990	99.0	7653	9	Q64742	ced-3 (C6434T) gene.
5	990	99.0	7653	9	Q64741	ced-3 (G6372A) gene.
6	990	99.0	7653	9	Q64738	ced-3 (C5940T) gene.
7	990	99.0	7653	9	Q64744	ced-3 (G6536A) gene.
8	990	99.0	7653	9	Q64743	ced-3 (C6485T) gene.
9	990	99.0	7653	9	Q64740	ced-3 (C6322T) gene.

10	990	99.0	7653	9	Q64737	ced-3 (G5757A) gene.	0.00e+00
11	990	99.0	7653	9	Q64745	ced-3 (C7020T) gene.	0.00e+00
12	990	99.0	7653	9	Q64739	ced-3 (G6297A) gene.	0.00e+00
13	988	98.8	7653	9	Q64736	ced-3 (G2487A) gene.	0.00e+00
14	144	14.4	2485	14	Q79970	Interleukin-1 beta co	9.23e-08
15	144	4.4	91	46	V44650	Mammalian DNA replica	2.53e-08
16	41	4.1	91	9	Q51746	Oligonucleotide probe	6.13e-07
17	39	3.9	91	9	Q51746	Oligonucleotide probe	4.97e-06
18	39	3.9	204	1	N81164	Base substituted E.co	4.97e-06
19	37	3.7	204	1	N81164	Base substituted E.co	3.93e-05
20	34	3.4	91	46	V44650	Mammalian DNA replica	8.24e-04
21	33	3.3	114	12	Q70467	Generic DNA sequence	2.24e-03
22	31	3.1	114	12	Q70468	Generic DNA sequence	1.60e-02
23	31	3.1	114	12	Q70468	Generic DNA sequence	1.60e-02
24	30	3.0	30	22	T31559	PCR primer pex1 for c	4.21e-02
25	30	3.0	30	9	Q54668	ced-3 gene primer Pex	4.21e-02
26	30	3.0	30	9	Q54667	ced-3 gene primer Pex	4.21e-02
27	30	3.0	30	24	T38199	Nematode Ced-3 gene p	4.21e-02
28	30	3.0	30	22	T31560	PCR primer Pex2 for c	4.21e-02
29	30	3.0	30	24	T38198	Nematode Ced-3 gene p	4.21e-02
30	30	3.0	114	12	Q70467	Generic DNA sequence	1.10e-01
31	29	2.9	114	12	Q70470	Generic DNA sequence	1.10e-01
32	29	2.9	114	12	Q70469	Generic DNA sequence	1.10e-01
33	29	2.9	3049	12	Q70890	Tomato spotted wilt v	2.82e-01
34	29	2.9	114	12	Q70466	Generic DNA sequence	2.82e-01
35	28	2.8	114	12	Q70470	Generic DNA sequence	2.82e-01
36	28	2.8	114	12	Q70465	Generic DNA sequence	2.82e-01
37	28	2.8	7653	9	Q54401	ced-3 gene.	2.82e-01
38	28	2.8	7653	9	Q54401	ced-3 (C6485T) gene.	2.82e-01
39	28	2.8	7653	9	Q64743	ced-3 (G6297A) gene.	2.82e-01
40	28	2.8	7653	9	Q64739	Nematode Ced-3 gene.	2.82e-01
41	28	2.8	7653	24	T38196	ced-3 (C6434T) gene.	2.82e-01
42	28	2.8	7653	9	Q64742	ced-3 gene.	2.82e-01
43	28	2.8	7653	9	Q54666	ced-3 (C5940T) gene.	2.82e-01
44	28	2.8	7653	9	Q64738	ced-3 (G6372A) gene.	2.82e-01
45	28	2.8	7653	9	Q64741	ced-3 (G6372A) gene.	2.82e-01

ALIGNMENTS

RESULT 1
ID 054401 standard; DNA: 7653 BP.
AC Q54401;
DT 01-JUL-1994 (first entry)
DE ced-3 gene.
KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
KW Interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
KW inflammatory response; nematode; diagnosis; myocardial infarction;
KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
KW pathogenic infection; hair loss; cancer; autoreactive antibody; ss.
OS Caenorhabditis elegans.
FH Key location/Qualifiers
FT repeat_unit 1356..1472
FT FT /*tag= a
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t repeat at 1490-1614"
FT FT 1490..1614
FT FT /*tag= b
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT FT 2232..2366
FT FT /*tag= c
FT FT /number= Exon_1
FT FT 2310
FT FT /*tag= d
FT FT /note= "C>T, from allele n1040"
FT FT 2367..2429
FT FT /*tag= e
FT FT /number= Intron_1
FT FT 2430..2575
FT FT /*tag= f
FT FT /number= Exon_2
FT FT 2487

C. elegans cdna yk170a12.5"
/codon_start=1
/protein_id="AAB42290.1"
/db_xref="PID:g1825673"
/db_xref="GI:1825673"
/translation="MSRFLSFCLLAALVSATSAOFTDCTTQDYAAITOCYVYMGY
NISMSTTVPEYWSFHFARLDLDADGLHIQPYVCQLGNALSNCLAPYSCMGNAIMQM
NAANTTEATDWDIDLAVTEYQGVNULMTTEYCMACFRDYOPIIDOCDAQAIDM
KNGMDPCAACQKDFNCQAAYVRNCCDMNAGVYICNVDAVGSKAIVNPACVNAGLVTCPA
PR"
complement(3730..6640)
/gene="F53E10.5"
complement(join(3730..3855,3955..4242,4314..4403,
4688..4728,5527..5694,5875..5931,6106..6195,6269..6332,
6435..6518,6572..6640))
/gene="F53E10.5"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB42291.1"
/db_xref="PID:g1825674"
/db_xref="GI:1825674"
/translation="MDNNSNVEFLTEELLEDLEIKTNWINEWQHIAVVKTNNERDADE
TINTEMKDPHDHNDNDNDPEIROPTEDRPSYTVNENHPDMPYORELPETIAFEE
NYFPSSRPNSTNGHYNDGNAROSTPQNOAEIAPISVPLVYAHANSOIGHFEMTLP
VFOEHQVMIHCQANQPEKPRHRRIKTELMLDIFIGTLQOEKARINEITEENKENIEH
LOKMTSTAERKLNDLKNKAPNDEKGMTIKGTGSRKTRAKOSKESKTLAYKIRLGE
VEIEKFQAISETLEQIAKNAQFVLENGIPANYFYERPLHTEDRSDVSHALDPTCSLL
PARRHIFGSNRAOEONAIASAIPQ"
complement(9953..10770)
/gene="F53E10.6"
complement(join(9953..10211,10265..10413,10478..10566,
10692..10770))
/gene="F53E10.6"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB42292.1"
/db_xref="PID:g1825675"
/db_xref="GI:1825675"
/translation="MSTKNRRLVLTEDSDDDNERHEMSSGGESGEEGPPSSVDGAGD
ADETVAFPAIERKKKVIKLLKKEQSLKSVKEYRIKALVLPDITDREKERNLRR
VATKGVOLFNNAVSDROKIMSVAVEKMTARERARORFDDGNFDSDFADSGYVGA
KKEVKGEDDDGEDMDIGEEDIDTGNYSDED"
12240..15811
/gene="F53E10.3"
join(12240..12318,12363..12562,12698..12767,15605..15662,
15709..15811)
/gene="F53E10.3"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB42289.1"
/db_xref="PID:g1825672"
/db_xref="GI:1825672"
/translation="MAETKWPSSKNSISISYEGSLTSVISHVVKFPIIGIRPRELKT
KLIFRASANSYKNDTFEKVSALYSFCSLFRHPLLDSTGDEVEMTHILKFFFEFK
FVEDAPYELFWHFPHQHFLFRNFQKIPIFKNLFRKRLSEALNENFPTPVKLFKLSTIF
FYEGMSYFS"
16384..21870
/gene="F53E10.2"
join(16384..16434,16487..16554,17235..17462,17579..17656,
17699..17860,18899..19194,19511..19653,20539..20673,
21691..21870)
/gene="F53E10.2"
/note="similar to acetylcholine receptor protein"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB42293.1"
/db_xref="PID:g1825676"
/db_xref="GI:1825676"
/translation="MISFQLLRIRIVLHFLMLGLAMAQOKLIKLFRDYHKELRPYKD
ESSGPTNVTQVLYEFKQIKQKHENDOIITLCWTEYWODEFITWNPDSFGNIKSLHVP
SEMIWKPDLVYNANANKENEMQTNVQIEHTGKISLFRALITDITCDLQMEKEPYD
QQTGIFMLASWSYDGSQIMLNTAEQTEAPLDIKNTTNLAILNHYPNMEWKLVDFRY
RNNLKYYDCPNPYPDISTYFFAIKRNPSTYLTLLIIPSAFITIVIGFTPHSSTGE

NTKESLGVTAALLSLAIIIMWSDKLPATNSVPLIGQYIGLIFIMFLATVCTTEFL
GIOMOGNAGRPIITRLRLSFLSIRMNNSFVQWFFGRELNTQESIKRLKYHNKTA
SLIEQESIDSSKLIDPAMYNNVVVYSSVQISIRDLTSGEHLKRIKKEWOMLAR"
23519..26175
/gene="F53E10.1"
join(23519..23587,23688..23798,23847..23935,24914..25060,
25114..25297,25708..25788,25837..26175)
/gene="F53E10.1"
/note="coded for by C. elegans cdna yk167d12.3; coded for
by C. elegans cdna yk167d12.3; coded for by C. elegans
cdna yk60a3.5; coded for by C. elegans cdna yk169e10.5"
/codon_start=1
/protein_id="AAB42288.1"
/db_xref="PID:g1825671"
/db_xref="GI:1825671"
/translation="MELDKDKVLLIGCLASSFIYHTARQIYSYSAKSREWPFIGI
VKGLHLYPIKSCKPDVYFAFKCTKLGPMMELEDVFLVVEESTGKFTARQPKLVN
VESYMTDGILEISVPGHPLKSLVDLKKVVENGRITRATLFDNLQODGYDCDEVAKLS
DYIEEPNRYLLFSKEGYTERTCIPDDQWNTVPVPRKDNSTGFTDLAFLIATEASLK
AVNEKLDKVTMRNFRPSIYIEGCAWDEDKWAEIRIGEHLCECAPCTRCVLTVDP
EKGEKSKEMQPLKLRERFLAPEGKMSKAHKDSPVFGVYAGTVNEXYIHIGTAYVKY
KPTVFE"
BASE COUNT 7844 a 4906 c 4801 g 8725 t
ORIGIN
Query Match 4.2%; Score 42; DB 21; Length 26276;
Best Local Similarity 92.0%; Pred. No. 5.20e-06;
Matches 46; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Db 20234 AAAGAGCGGAGCGTAATTTGCAACCTGCGGACGCTTTTCTTCC 20283
OY 2619 AAAGAGCGGATCGTAATTTTGCACACCGGACGCTTTTCTTCC 2668
RESULT 14
LOCUS AC006710 145920 bp DNA HTG 23-FEB-1999
DEFINITION Caenorhabditis elegans clone Y119C1AA, WORKING DRAFT SEQUENCE, 1
unordered pieces.
AC006710
AC006710
AC006710.1 GI:4263228
HTG: HTGS_PHASE1.
KEYWORDS
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 145920)
Waterston,R.H.
The sequence of Caenorhabditis elegans clone
Unpublished
2 (bases 1 to 145920)
Waterston,R.H.
Direct Submission
Submitted (23-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 145920: contig of 145920 bp in length.
FEATURES
source
1..145920
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y119C1AA"
BASE COUNT 46442 a 26006 c 25762 g 47710 t
ORIGIN

Note: remainder of annotations omitted.

[illegible]

RESULT	12				
LOCUS	CEY46G5	315330 bp	DNA	HTG	13-MAR-1999
DEFINITION	Caenorhabditis elegans DNA	*** SEQUENCING IN PROGRESS	***	from	
ACCESSION	clone Y46G5, WORKING DRAFT SEQUENCE.				
	703957				

REFERENCE	1 (bases 1 to 315330)	
AUTHORS	wallis, J.	
TITLE	Direct Submission	
JOURNAL	Submitted (13-MAR-1999)	Nematode Sequencing Project
		Sanger Centre

COMMENT

On Mar 23, 1999 this sequence version replaced gi:4481799.

* NOTE: This is a 'working draft' sequence.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

```
/organism="Caenorhabditis elegans"  
/db_xref="taxon:6239"
```

Query Match	4.3%;	Score 43;	DB 19;	Length 315330;
Best Local	Similarity 75.3%;	Pred. No. 1.75e-06;		
Matches	64;	Conservative	0;	Mismatches 21;
			Indels	0;
			Gaps	0;

QY 2680 ATTATGCACTTTCCTCCAAATTTTGAAGTGAATATATTTTATTACTGAAAGCTCGAGTG 2739

RESULT 13

ACCESSION U88177

VERSION U88177.1 GI:1825670

SOURCE *Caenorhabditis elegans* strain=Bristol N2

ORGANISM
Caenorhabditis elegans
Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea,
Rhabditidae; Peloderinae; Caenorhabditis.

TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*

MEDLINE

AUTHORS

REFERENCE

TITLE

Submitted (03-FEB-1997)
Submitted by:

e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems such as compressions and repeats; all regions were covered by sequence from more than one subclone

The 5' cosmid is C29G2, 300 bp overlap; 3' cosmid is W02H5. Actual start of this cosmid is at base position 1 of CELF53E10. This cosmid lies in an unanchored cluster the orientation of which is unknown.

FEATURES

```
/organism="Caenorhabditis elegans"
```

```
/scran- disc01 NZ
/db xref="taxon:6239
```

```
/chromosome="Y"
```

```
/clone="F53E10"
```

```
/gene="F53E10.4"
```

John(304: .539,1636. .1787,1831. .1945,2563. .2722)
/2000-11-23 10:41"

/note="coded for by C. elegans cDNA yk65d4.5; coded for by

The true left end of clone F35C5 is at 1 in this sequence. The true right end of clone F35C5 is at 8095 in sequence 281556.
The true left end of clone F58G1 is at 33064 in this sequence. The start of this sequence overlaps with the end of sequence 292852.
The end of this sequence (33064..33164) overlaps with the start of sequence 281556.

```
FEATURES
source
    location/Qualifiers
        1..33164
        /organism="Caenorhabditis elegans"
        /db_xref="taxon:6239"
        /chromosome="II"
        /clone="F35C5"
        2293..3419
        /gene="F35C5.1"
        join(2293..2396,2761..2934,3207..3326,3374..3419)
        /gene="F35C5.1"
        /codon_start=1
        /protein_id="CAB03050.1"
        /db_xref="PID:e1346293"
        /db_xref="PID:g3876676"
        /db_xref="GI:3876676"
        /db_xref="SPTREMBL:O45437"
        /translation="MCLHLMSVLFLEWPLILVLSLVVCVSKKSQKRIDS LKVP RPQASSMSSSGSKSYRKSGNOVSRGSDTSDTSAKTELIEMPLGKAARDEPKHNILT PAAEQKTGDSBPSKRVQFAHPLKRVKRRATMPEDADEKMTYA"
        6754..10069
        /gene="F35C5.2"
        join(6754..6993,7040..7185,7766..7841,7890..8089,8366..8552,9216..9256,9358..9410,9879..10069)
        /gene="F35C5.2"
        /codon_start=1
        /protein_id="CAB03051.1"
        /db_xref="PID:e1346294"
        /db_xref="PID:g3876677"
        /db_xref="GI:3876677"
        /translation="MSEECPFKASDDQIILQTSWILRLNIYCTFLSIGCFVGVAYC IRFMRKHPIFESSTAII LYLSLAFYHDVAHVL SQWVMYRSIFYADPCNILEFSD DCLSLGRSIVFGISGLIFIHSLADGLLETFTPMFYKQRMCGIILAVLMISVIG VQFYLLPIGESKDYLPSCQFFRKQDAPRANFLISSLVLTITLNMVNTLLFVNKMH SKTRFEDVEFOYQORSEAMMTSESISLVIAQISALGIAGSNLFRQAKDEIPVLYXS NLIWVYAI SGFIEHFLKFGIAAHPYMNKSTNGQIAGNLLRRNPPTSHHILHSIC KTA PPA DSHSHKRPQGEPSQNAGAAALGIIVLWFLCEILK"
        10256..11938
        /gene="F35C5.3"
        join(10256..10321,10369..10455,10501..10557,10605..11249,11302..11355,11411..11455,11506..11547,11595..11630,11678..11710,11762..11827,11885..11938)
        /gene="F35C5.3"
        /note="cDNA EST yk207f10.3 comes from this gene; cDNA EST yk207f10.5 comes from this gene"
        /codon_start=1
        /protein_id="CAB03052.1"
        /db_xref="PID:e1346295"
        /db_xref="PID:g3876678"
        /db_xref="GI:3876678"
        /db_xref="SPTREMBL:O45438"
        /translation="MKLSLVFLFVAFKASADTAPVDENKIVTFVEELMDAINQLIL EOMGAIRPSTPLQSIPIPTADTADTAETPTKAHETVAQEATSEGAETHETV GQDTTVRSGESQETTVSQGTIVSQETTVSQETTVSQETTVSQETTVSQETTV SQESTVSQETTVSQDTVSQETSVSHETNVPDITVSQESTVSQDTVSQETTLISQET TVSQETTVSQDTVSQETTVSQDTVSQDTTVVPOETTVETATKASEATEENG TTA VTPGESDPDTEATPTGGSTVGESPECDTTEASEPTGHTDOAETVTPASQATAS ATIIIGVDQTGTNLEATTEQICKHPTKDKSEEEKTDENKDEKTRAGTNVHRPERDI LK"
        12108..13502
        /gene="F35C5.4"
        join(12108..12128,12366..12551,12916..12993,13048..13095,13141..13185,13241..13279,13338..13379,13443..13502)
        /gene="F35C5.4"
        /codon_start=1
        /protein_id="CAB03053.1"
        /db_xref="PID:e1346296"
        /db_xref="SPTREMBL:O45439"
        /translation="MGASTATVAIIQSEPTSSSSSPSSRPEDCATVGNFTQSSAV SESQTSERTSVSPDSTPVRVTTSTILSTVPTSDISGVSTADPTKATLPITYP PPTTTINLAPITYPPILTITISLEPIFEQVPVNLPIFISQNPVPSNEKNPLSPIF NDVPSAYQPLIW"
        complement(14630..16576)
        /gene="F35C5.5a"
        complement(14630..16576)
        /gene="F35C5.5b"
        complement(join(14630..14809,14862..14974,15032..15229,15289..15584,15632..15735,16263..16376,16418..16576))
        /gene="F35C5.5a"
        /note="similar to lectin C-type domain short and long forms, von Willebrand factor type A domain; cDNA EST yk204f1.3 comes from this gene"
        /codon_start=1
        /protein_id="CAB03054.1"
        /db_xref="PID:e1346297"
        /db_xref="PID:g3876680"
        /db_xref="GI:3876680"
        /db_xref="SPTREMBL:O45440"
        /translation="MLIGLLFLLSVYTGQVPTDRNCGHDLANLMDVVVVDNSA PMTQEGLEVAQIVTVFGAGTRIGTYIDKRTTRVGLVYNTETATIQADINRFQSPD DLFTVFQILPNDLSTSEDFVLAKGIGAEQLLAAGRKNNTKRNKQMVIVASAYND EGEEPRPIAERLKASGVSIATVAFQDGTDEEMIKLIGEIATPGNFNEDENLVKEI QTAMIQTNVCYNLMTQYRTEFEDPHSKFGVCLRPVALTAGWTPAKACQMIOQGY MTEYDQKHDPYVKLVNDTSFPEPYVYHIGLSYNGGYAQDPVGHALVPLNDSLW NPGFPTQSTTTAVLNQOSSSAFRVGMQNVNQTYSERYCEVAACTSTYCE"
        complement(join(14630..14815,14862..14974,15032..15229,15289..15584,15632..15735,16263..16376,16418..16576))
        /gene="F35C5.5a"
        /note="similar to lectin C-type domain short and long forms, von Willebrand factor type A domain; cDNA EST yk367h4.3 comes from this gene; cDNA EST yk367h4.5 comes from this gene; cDNA EST yk363d9.3 comes from this gene; cDNA EST yk363d9.5 comes from this gene; cDNA EST yk204f1.5 comes from this gene"
        /codon_start=1
        /protein_id="CAB03055.1"
        /db_xref="PID:e1346298"
        /db_xref="PID:g3876681"
        /db_xref="GI:3876681"
        /db_xref="SPTREMBL:O45441"
        /translation="MLIGLLFLLSVYTGQVPTDRNCGHDLANLMDVVVVDNSA PMTQEGLEVAQIVTVFGAGTRIGTYIDKRTTRVGLVYNTETATIQADINRFQSPD DLFTVFQILPNDLSTSEDFVLAKGIGAEQLLAAGRKNNTKRNKQMVIVASAYND EGEEPRPIAERLKASGVSIATVAFQDGTDEEMIKLIGEIATPGNFNEDENLVKEI QTAMIQTNVCYNLMTQYRTEFEDPHSKFGVCLRPVALTAGWTPAKACQMIOQGY MTEYDQKHDPYVKLVNDTSFPEPYVYHIGLSYNGGYAQDPVGHALVPLNDSLW LMPGFPTQSTTTAVLNQOSSSAFRVGMQNVNQTYSERYCEVAACTSTYCE"
        17118..18865
        /gene="F35C5.6"
        join(17118..17665,17998..18194,18243..18553,18686..18865)
        /gene="F35C5.6"
        /note="similar to lectin C-type domain short and long forms, von Willebrand factor type A domain; cDNA EST CEMS87F comes from this gene; cDNA EST yk279b9.5 comes from this gene; cDNA EST EMBL:D69169 comes from this gene; cDNA EST EMBL:D69482 comes from this gene; cDNA EST EMBL:D65966 comes from this gene; cDNA EST EMBL:D69586 comes from this gene; cDNA EST EMBL:D69649 comes from this gene; cDNA EST EMBL:D66470 comes from this gene; cDNA EST EMBL:D70265 comes from this gene; cDNA EST EMBL:C11927 comes from this gene; cDNA EST EMBL:C10473 comes from this gene; cDNA EST EMBL:C12088 comes from this gene; cDNA EST EMBL:C11616 comes from this gene; cDNA EST yk207c12.3 comes from this gene; cDNA EST yk207c12.5 comes from this gene; cDNA EST yk234d2.3 comes from this gene; cDNA EST yk235f6.3 comes from this gene; cDNA EST yk235f6.5 comes from this gene; cDNA EST yk261g11.3 comes from this gene; cDNA EST
```

```
/db_xref="PID:g3876679"
/db_xref="GI:3876679"
/db_xref="SPTREMBL:O45439"
/translation="MGASTATVAIIQSEPTSSSSSPSSRPEDCATVGNFTQSSAV SESQTSERTSVSPDSTPVRVTTSTILSTVPTSDISGVSTADPTKATLPITYP PPTTTINLAPITYPPILTITISLEPIFEQVPVNLPIFISQNPVPSNEKNPLSPIF NDVPSAYQPLIW"
complement(14630..16576)
/gene="F35C5.5a"
complement(14630..16576)
/gene="F35C5.5b"
complement(join(14630..14809,14862..14974,15032..15229,15289..15584,15632..15735,16263..16376,16418..16576))
/gene="F35C5.5a"
/note="similar to lectin C-type domain short and long forms, von Willebrand factor type A domain; cDNA EST yk204f1.3 comes from this gene"
/codon_start=1
/protein_id="CAB03054.1"
/db_xref="PID:e1346297"
/db_xref="PID:g3876680"
/db_xref="GI:3876680"
/db_xref="SPTREMBL:O45440"
/translation="MLIGLLFLLSVYTGQVPTDRNCGHDLANLMDVVVVDNSA PMTQEGLEVAQIVTVFGAGTRIGTYIDKRTTRVGLVYNTETATIQADINRFQSPD DLFTVFQILPNDLSTSEDFVLAKGIGAEQLLAAGRKNNTKRNKQMVIVASAYND EGEEPRPIAERLKASGVSIATVAFQDGTDEEMIKLIGEIATPGNFNEDENLVKEI QTAMIQTNVCYNLMTQYRTEFEDPHSKFGVCLRPVALTAGWTPAKACQMIOQGY MTEYDQKHDPYVKLVNDTSFPEPYVYHIGLSYNGGYAQDPVGHALVPLNDSLW NPGFPTQSTTTAVLNQOSSSAFRVGMQNVNQTYSERYCEVAACTSTYCE"
complement(join(14630..14815,14862..14974,15032..15229,15289..15584,15632..15735,16263..16376,16418..16576))
/gene="F35C5.5a"
/note="similar to lectin C-type domain short and long forms, von Willebrand factor type A domain; cDNA EST yk367h4.3 comes from this gene; cDNA EST yk367h4.5 comes from this gene; cDNA EST yk363d9.3 comes from this gene; cDNA EST yk363d9.5 comes from this gene; cDNA EST yk204f1.5 comes from this gene"
/codon_start=1
/protein_id="CAB03055.1"
/db_xref="PID:e1346298"
/db_xref="PID:g3876681"
/db_xref="GI:3876681"
/db_xref="SPTREMBL:O45441"
/translation="MLIGLLFLLSVYTGQVPTDRNCGHDLANLMDVVVVDNSA PMTQEGLEVAQIVTVFGAGTRIGTYIDKRTTRVGLVYNTETATIQADINRFQSPD DLFTVFQILPNDLSTSEDFVLAKGIGAEQLLAAGRKNNTKRNKQMVIVASAYND EGEEPRPIAERLKASGVSIATVAFQDGTDEEMIKLIGEIATPGNFNEDENLVKEI QTAMIQTNVCYNLMTQYRTEFEDPHSKFGVCLRPVALTAGWTPAKACQMIOQGY MTEYDQKHDPYVKLVNDTSFPEPYVYHIGLSYNGGYAQDPVGHALVPLNDSLW LMPGFPTQSTTTAVLNQOSSSAFRVGMQNVNQTYSERYCEVAACTSTYCE"
17118..18865
/gene="F35C5.6"
join(17118..17665,17998..18194,18243..18553,18686..18865)
/gene="F35C5.6"
/note="similar to lectin C-type domain short and long forms, von Willebrand factor type A domain; cDNA EST CEMS87F comes from this gene; cDNA EST yk279b9.5 comes from this gene; cDNA EST EMBL:D69169 comes from this gene; cDNA EST EMBL:D69482 comes from this gene; cDNA EST EMBL:D65966 comes from this gene; cDNA EST EMBL:D69586 comes from this gene; cDNA EST EMBL:D69649 comes from this gene; cDNA EST EMBL:D66470 comes from this gene; cDNA EST EMBL:D70265 comes from this gene; cDNA EST EMBL:C11927 comes from this gene; cDNA EST EMBL:C10473 comes from this gene; cDNA EST EMBL:C12088 comes from this gene; cDNA EST EMBL:C11616 comes from this gene; cDNA EST yk207c12.3 comes from this gene; cDNA EST yk207c12.5 comes from this gene; cDNA EST yk234d2.3 comes from this gene; cDNA EST yk235f6.3 comes from this gene; cDNA EST yk235f6.5 comes from this gene; cDNA EST yk261g11.3 comes from this gene; cDNA EST
```


MEDLINE	97132596
REFERENCE	2 (bases 1 to 10772)
AUTHORS	Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.
TITLE	Structure of the cysteine proteinase (Cp1) gene of <i>Drosophila melanogaster</i> and associated mutational effects
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 10772)
AUTHORS	Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.
TITLE	Direct Submission
JOURNAL	Submitted (30-JUN-1997) School of Biological Sciences, University of Sydney, Biology A12, Sydney University, NSW 2006, Australia
FEATURES	Location/Qualifiers
source	1..10772
MRNA	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	join(872..1000,2310..2426,6476..6690,6751..7707)
	/gene="Cp1"
gene	/product="cysteine protease"
	872..7707
	/gene="Cp1"
exon	872..1000
	/gene="Cp1"
	/number=1
intron	1001..2309
	/gene="Cp1"
exon	2310..2426
	/gene="Cp1"
	/number=2
CDS	join(2328..2426,6476..6690,6751..7462)
	/gene="Cp1"
	/codon_start=1
	/product="cysteine proteinase-1"
	/protein_id="AAB65749.1"
	/db_xref="PID:g2305221"
	/db_xref="GI:2305221"
	/translation="MRTAVLLPLLALAVAVSFADVYMEWHTEKLEHRKNYODET EERFRKIENENKHKIAKHNRFAEGKVSFKLVANKYADLLHFEFRLMNGFNVTIAK QLRAADESEFKVTFISPAHVTLPKSVDMRTGAVTAVKDQHGSCWAFSSGTALGEG HFRKSGVLVSLSEQLNVDSTKYGNNGCGLMDNAFRYIKDNGGIDTEKSPYEAI DSCHEFNKGTGATDRGFTDIPQDEKMAEAVATGVPVSAIDASHESFOFYSEGVYN EPQCDQNLDHGYLVVGFGTDESGEDYWLVRNSWGTWGDKGFIKMLRNKENQCIAS ASSYPLV"
intron	2427..6475
	/gene="Cp1"
	4546..4553
misc_feature	/gene="Cp1"
	/note="insertion site of P{Casper}(50C)"
	/citation=[1]
exon	6476..6690
	/gene="Cp1"
	/number=3
intron	6691..6750
	/gene="Cp1"
exon	6751..7707
	/gene="Cp1"
	/number=4
MRNA	join(<8110..9300,9370..>9532)
	/product="phenylalanyl tRNA synthetase"
	join(8110..9300,9370..>9532)
	/note="potential orf"
	/codon_start=1
	/product="phenylalanyl tRNA synthetase"
	/protein_id="AAB65750.1"
	/db_xref="PID:g2305222"
	/db_xref="GI:2305222"
	/translation="MLTLRVQGARHMLKSTRCLASSAAPAKSPSPPOLEYSGSTYA TDGWTNTPKILSYVGANKHLQTDHPLISIRQRLVNYFYGAYRNQGNPLFSYDOMNNTLVG PVTVVQNFNDLLIPADHVSROKSDCYINQOHLRAHTTAHVELISGGLDNFLVVG EYVRDEIDSTHYPVFHQADAVRLVTKDKLFERNPGLLEFETWSGTLADPKLILPH SSWTKPNSPATRAVKLMEHEMKHVLVGLTKDLFGPRIKRYMVDYTFPTOPSWLEII YFKDNMLEVLGGGIMRHEILQRSVHQSIGYAFGVLERLAMLFDIPDIRLFWNSDS GFLSQSEKDLNLPKYKPISHYPQCTNDLSFWLPQDIEVDAGSPNDFYDLVRSVAG DMEQISLVDKFKRHKDKTGKSSVCFRIVRYRHMERTLLQAEVNEIHKQIASAVDSFNVO
CDS	

BASE COUNT	2929 a	2357 c	2282 g	3046 t	158 others
ORIGIN	IX"				
Query Match	4.3%; Score 43; DB 21; Length 10772;				
Best local Similarity	19.9%; Pred. No. 1.75e-06;				
Matches	32;	Conservative	75;	Mismatches	52; Indels 2; Gaps 2;

Db	1677	AGARAGTTTCTTAACRWWYAWTTTMKKMMWTTWKAMMKTYRTWMWKMYSRTTTT	1736
		: : : : : : : :: :: : : :: : : :	
Cp	2011	AGTAAATCGGAAAATTTCACCCTCTGGAATTTTTTAATCTTGATTTCTGGCGA-ATTTT	1953
Db	1737	SAMMMWYTSTWTXYWAYAWMKMWTTRTWARMASWARWKWTSAAAAYSAW-RKMWK	1795
		: : : : :: : : : : : : : : : : : : : : : :	
Cp	1952	GGGGTTTTTTTTTGTTAAATTTGTTTGAATCAGTTCTCTTTATTTTCGCAACATPAA	1893
Db	1796	WWAYRAMKKTMMMAWKRKAAMWTTWRWYMTTTAAMRAA	1836
		: : : : : : : : : : : : : : : : : : :	
Cp	1892	ACACAATTTTCAATGAGAATCGTTATTTTAAATTCAGAAA	1852

RESULT	11
LOCUS	CEF35C5 33164 bp DNA INV 23-NOV-1998
DEFINITION	Caenorhabditis elegans cosmid F35C5, complete sequence.
ACCESSION	Z81076
NID	g1627897
VERSION	Z81076.1 GI:1627897
KEYWORDS	HTG.
SOURCE	Caenorhabditis elegans.
ORGANISM	Caenorhabditis elegans.
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS	Sims,M.
JOURNAL	Direct Submission
REFERENCE	Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu
AUTHORS	2 (bases 1 to 33164)
	Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Shownkeen,R., Smaildon,N., Smith,A., Sonhammer,E., Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Spratt,J. and Wohlman,P.
TITLE	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL MEDLINE COMMENT	Nature 368 (6466), 32-38 (1994) 94150718 Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information. For a graphical representation of this sequence and its analysis see:- http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence &object=F35C5 Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. IMPORTANT: This sequence is not the entire insert of clone F35C5. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

/evidence=not_experimental
/protein_id="AAC19217.1"
/db_xref="PID:g3193188"
/db_xref="GI:3193188"
/translation="MYEDQYCVGNFLFGTDFREDGLEPITLASHATSDMLTLEKM
TSMWDGPISVGIFIDFSSQALEYLAHVHRCDEEFKRMTHFAIKQSAFOOTCPKIQ
IPASDRTCKWFRADQSYLRSHLSGPFQLYPSNLMRLARQAKSDIHFIMDADMTVSE
GFARKLKVANEMIDGSKKVLAIREFESVNGTYLPTHELFQSMAYSNQYEWEOV
ILHRNDPYNAAFSPSRKVMHSLIALCALCRAGTFHPVSHVFDVHEGIKHTNTIYSKAT
IAHQEAYAMDIAGARVYREMDKYPDTLDCGCFKMY"
19348..23506
/gene="F22F7.2"
join(19348..19506,19621..19740,20773..21033,21767..22174,
22868..23082,23401..23506)
/gene="F22F7.2"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC19222.1"
/db_xref="PID:g3193193"
/db_xref="GI:3193193"
/translation="MEARYDVVIYATGYTGKIFETLISCGKFDNYSIAAGRSEKK
LEEVLEKLEKSTSLKTRIGLVCDSTNEESMGKARRAKLIYNAVGPRLHGEAVR
KAAVENGANOIDVAGEPEWIERMEAKYGMARNNNVIVSACGWSIPADFGVKLKR
YFDGLQRIDAFLOLHFPGSGYSFSAISQALLGFGNAPDLGSLRHSIMPKIDHL
APNKRHFHWKIEKSGSEGWAMPFLGADKSIVTRSQYFDYVMNNVYPVPTPTRES
RWNALLLAALMGILKTFSGYTMRFIMRPELCSGLFSGSGSEQMKFATFTYQF
YGYGGRGDPDRPREPKKLVTCTGPDVGVIATSGCVLSSVLYFLKEKEKLPKSGGVY
TTAYAFENTRLIDFLDFGIKFDIEMPSKL"
24016..24017
/note="SL1 trans-splice site; see EST cm1263"
24018..26348
/gene="F22F7.1"
join(24018..24183,24228..24334,24381..24801,25685..26165,
26243..26348)
/gene="F22F7.1"
/note="coded for by C. elegans cDNA yk139b8.3; coded for
by C. elegans cDNA cm14g4; coded for by C. elegans cDNA
yk59a9.5; coded for by C. elegans cDNA yk139b8.5; coded
for by C. elegans cDNA cm1263; coded for by C. elegans
cDNA cm15e5"
/codon_start=1
/protein_id="AAC19221.1"
/db_xref="PID:g3193192"
/db_xref="GI:3193192"
/translation="MSRYDVVYVYGSGFTGAYVEYLVNSEQFEGLSFAVAGRSEKKL
REVLNISOKTGKDVSNAAVIVADSDERSLNEMARQANVINAVPYRLYGEAVVKA
AVENGASHVDISGEPAWIEKMOQKYSKQAKQGVYVVSACGWSIPADLVNFKKNF
HGLDNHVESFVOLLTGPSGYSFNAGTYQLILGLSGAATDKLGAVRKEIMPEKIVRGA
VKLPKRPFLWEIKELNGVAVPFGADKSIINRSQYDATSRQVRPIHMETYIRLSS
OFYGLIGLWIMFLSIFVKYPFTRRIQQYPPDQCSFYMFKNPSGARQMAEASFVYMF
FGYKKTLPDQOHEGKINRKVLATCGPDAGVIATSGCVLSAALLIRDKDNLPRKD
GGVYTTAAAFGNSKIYDYLASFQITQLESEYDL"
complement(28572..30790)
/gene="F22F7.7"
complement(join(28572..28680,29262..29452,29953..30069,
30620..30790))
/gene="F22F7.7"
/note="contains similarity to E. coli cation transport
protein (GB:I28709)"

Query Match 4.4%; Score 44; DB 21; Length 41564;
Best Local Similarity 75.6%; Pred. No. 5.82e-07;
Matches 65; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Db 24994 GAAATATAGTGTCAAGTATAAAATGGCTCGATTTCTGAAATATAATTTTTCGA 25053
|||||
Cp 2766 GAAATATAGTGTAAAAATAATAATCACTCGAGCTTCAAGTAAATAATATATTTC 2707
|||||
Db 25054 GCTGAACAATTTGGGAAAGTGACAA 25079
|||||
Cp 2706 CTTCAATAATTGGGAAAGTGCAITAA 2681
|||||

RESULT 9
LOCUS CEY53F4 495882 bp DNA HTG 02-MAR-1999
DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
clone Y53F4, WORKING DRAFT SEQUENCE.
ACCESSION Z92860
NID 94464195
VERSION Z92860.4 GI:4464195
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 495882)
REFERENCE
AUTHORS Smye,R.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-1999) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@anger.ac.uk or rwenematode.wustl.edu
On Mar 22, 1999 this sequence version replaced gi:4455382.
COMMENT
Order of segments is not known; 800 n's separate segments.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
source location/Qualifiers
1..495882
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="II"
/clone="Y53F4"
BASE COUNT 153420 a 87589 c 84294 g 153691 t 16888 others
ORIGIN
Query Match 4.4%; Score 44; DB 19; Length 110000;
Best Local Similarity 76.8%; Pred. No. 5.82e-07;
Matches 63; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Db 56849 GCATTTTCACAAACATTTCAACTAAATAATGTTATTTCAATAATCAAGTATTTT 56908
|||||
QY 2685 GCACCTTCCCAATATTTGAAGCAAAATATATTTTAACTGAAAGCTCGAGTATTA 2744
|||||
Db 56909 CTATTTTTCACACTAATTCCTC 56930
|||||
QY 2745 TTATTTTAAACACTAATTTTC 2766
|||||

RESULT 10
LOCUS AF012089 10772 bp DNA INV 05-AUG-1997
DEFINITION Drosophila melanogaster cysteine proteinase-1 (Cp1) gene, complete
cds, and phenylalanyl tRNA synthetase gene, partial cds.
ACCESSION AF012089
NID 92305220
VERSION AF012089.1 GI:2305220
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; mitochondrial eukaryotes; Metazoa; Arthropoda;
Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS Gray,Y.H., Tanaka,M.M. and Sved,J.A.
TITLE P-element-induced recombination in Drosophila melanogaster: hybrid
element insertion
JOURNAL Genetics 144 (4), 1601-1610 (1996)

ORGANISM
Caenorhabditis elegans
Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
Secernentea; Rhabdilia; Rhabdilitida; Rhabdilitina; Rhabditoidea;
Rhabdilitidae; Caenorhabditis.
1 (bases 1 to 35713)
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,
Saunders, D., Showkhen, R., Smaison, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Woldman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL
Nature 368 (6466), 32-38 (1994)
MEDLINE
94150718
2 (bases 1 to 35713)
Leimbach, D.
TITLE
The sequence of C. elegans cosmid F22D3
JOURNAL
Unpublished (1995)
REFERENCE
3 (bases 1 to 35713)
AUTHORS
Waterston, R.
TITLE
Direct Submission
JOURNAL
Submitted (12-JUN-1995) Robert Waterston
COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: rtw@nemacode.wustl.edu and jes@sanger.ac.uk

NEIGHBORING COSMID INFORMATION:
The 5' cosmid is B0252, 200 bp overlap; 3' cosmid is C15F1. Actual
start of this cosmid is at base position 1 of CELF22D3; actual end
is at base position 35713 of CELF22D3.

NOTES:
Coding sequences below are predicted from computer analysis, using
the program Genefinder (P. Green and L. Hillier, ms in
preparation).
FEATURES
source
Location/Qualifiers
1. .35713
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/map="II"
complement(join(108. .263,313. .385,428. .504,878. .955))
/gene="F22D3.4"
complement(join(108. .263,313. .385,428. .504,878. .955))
/gene="F22D3.4"
complement(join(108. .263,313. .385,428. .504,878. .955))
/gene="F22D3.4"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA68395.1"
/db_xref="PID:g861378"
/db_xref="GI:861378"
/translation="MNLVLYLVSSTFLCLAAEITYKTNNGKGEKFGSKDGPEAEAVEN
SDIVEDQNAOSSARSLGKPYEEDGFLVIENANVYSGVYFCFDEDSVASQREFFYL
IPLVPHIDSKNSKIMELSKFY"
join(1168. .1258,2416. .2438,2944. .2965,3774. .3794,
4215. .4251,4310. .4410,5631. .5710,5761. .5833,5919. .6003,
6131. .6179,6619. .6796,7216. .7282,7331. .7405,7451. .7583,
8018. .8140,9089. .9163)
/gene="F22D3.3"
join(1168. .1258,2416. .2438,2944. .2965,3774. .3794,
4215. .4251,4310. .4410,5631. .5710,5761. .5833,5919. .6003,
6131. .6179,6619. .6796,7216. .7282,7331. .7405,7451. .7583,
8018. .8140,9089. .9163)
/gene="F22D3.3"
CDS
join(1168. .1258,2416. .2438,2944. .2965,3774. .3794,
4215. .4251,4310. .4410,5631. .5710,5761. .5833,5919. .6003,
6131. .6179,6619. .6796,7216. .7282,7331. .7405,7451. .7583,
8018. .8140,9089. .9163)
/gene="F22D3.3"

/codon_start=1
/evidence=not_experimental
/protein_id="AAA68396.1"
/db_xref="PID:g861379"
/db_xref="GI:861379"
/translation="MFASEHTLSVRHKMDVIRVATTLNGKRRARSFWFLMIRTLIIIV
LDIHPIPDFFKLDQKICEIIQMVAIALCHFCENHGPRVMTCPMRDVEDKASC
TASTSLGSTSSPGIIVLYLFDGECVCKEKTNTEDGTSYPHYGNVRAHPHEMD
LAFQMIKQKRATSLQSSKRYEQNCSLDIISVALQDRYERVKNACLRSLSCEVS
AASKNPVOPPPPLSRSTPRKIHPIVIEEPLHGTIRIPSSTVLOEPEDGAFQEP
AAEEADGVIFGDTNGCFAYIFRLADAKARGYRLFLSVISNDITFLTNFEYFK
TTLGKINDLOKLAQDVFSREIDTGEKLSEDIHKPEFKLVGKMPSWYRRKIAIDTD
RNLTVTANNDIWWQLHR"
complement(join(13589. .13636,13700. .13751,13824. .13879,
13922. .14017))
/gene="F22D3.5"
complement(join(13589. .13636,13700. .13751,13824. .13879,
13922. .14017))
/gene="F22D3.5"
complement(join(13589. .13636,13700. .13751,13824. .13879,
13922. .14017))
/gene="F22D3.5"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA68397.1"
/db_xref="PID:g861380"
/db_xref="GI:861380"
/translation="MAAYSRTTFSENFESQFNCFTRLFSCCKHFNLMSTPARKPRK
ETENCKFVSSKSLADFGDVFSRQFERKNMSKEKTS"
join(19428. .19559,19742. .19852,21134. .21255,21436. .21802,
21850. .21951)
/gene="F22D3.2"
join(19428. .19559,19742. .19852,21134. .21255,21436. .21802,
21850. .21951)
/gene="F22D3.2"
join(19428. .19559,19742. .19852,21134. .21255,21436. .21802,
21850. .21951)
/gene="F22D3.2"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA68398.1"
/db_xref="PID:g861381"
/db_xref="GI:861381"
/translation="MPLIETQNIIPPSFMRVHEKNEALSALATEPDDFECENETIKGK
VSMQOLANLKLIATQNTCSPEGLDLIRHIVTGGOVVESEKEKRCROFLAISNL
LPIGCIKLATYKEQYLPEYSSRFNLIGPHNMDIPEVSDVMVFRVSPRDSIVDIEQ
MSLVNVCVIEVRPRPENDSGNPLPRIVNRYRDLSDSDVTDLIETALRTREGWMLK
AKICYQMTKQLSHPOVFELITGCGVEDQQVSVYTAGLSNAYKLHVLTISQOSTSTA
R"
complement(join(22981. .23313,23359. .23447,23739. .23888,
23962. .24058,24106. .24220,24428. .24501,24550. .24662,
24711. .24810,25106. .25329,25379. .25539,25582. .25751,
26050. .26131,26185. .26306))
/gene="F22D3.6"
complement(join(22981. .23313,23359. .23447,23739. .23888,
23962. .24058,24106. .24220,24428. .24501,24550. .24662,
24711. .24810,25106. .25329,25379. .25539,25582. .25751,
26050. .26131,26185. .26306))
/gene="F22D3.6"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA68399.1"
/db_xref="PID:g861382"
/db_xref="GI:861382"
/translation="MNTNLAEPLVRIKILSELOKDNIMIKIVEFDEDPPIYMKSTE
IESFLKREKCEQVLRKMGNGQTSLYGYPDEQOYTKIEIDGNIKLECKAOGFPCP
EIKMFTKSGSKPEVHIGRVYTLIRCKCSNEHQYKCAVAKNEIREGSPYSEIYRKAGKQFS
SVIESEYDVNTSCIRDELCECKRYEMGRLSQILAEQEKPEKNKSPVRPNLDITLR
AADKVALIMNSQSVYHLPRLPHCDATQTLADALQKMYKTVTLADLTLDREYFIRY
YOKLIGNVYAVFYVGHGFVNGOCYLIGVDAPADAPADAPQHSMSMDWLSIFRKTTP
DLNLLLDYCRKFVPIYDAISAVEYSEQFKFRAHRKNVYGYSTSGVGAYEVKGEY
NGVEMKYLKNHVLLEISVIDMLNKVLDIGDDQKVCDOVPEIRSTLTHTPRSLADPLI
FDGHTASEDNHTIHWRLMHELPTPATIRFETQQLVATITWFOCGNFTNKVYVFAISID
FRPCQEDTDMGENEELSINALNHRFAVEFPEELHCSDVRENDDEGVSMLWILSLQ
KIKKEGLTCEVHLRHVDDEPTEIEMKNVDIGHLITRIKLO"
join(27077. .27114,27160. .27471,27520. .27622,27751. .28066,
28189. .28675,28796. .29036,29170. .29256,29386. .29736)
/gene="F22D3.1"
join(27077. .27114,27160. .27471,27520. .27622,27751. .28066,
28189. .28675,28796. .29036,29170. .29256,29386. .29736)
/gene="F22D3.1"

6498. .6602,6658. .6748,6791. .7021)
/gene="C16A11.5"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC26261.1"
/db_xref="PID:g3319386"
/db_xref="GI:3319386"
/translation="MCSLNLITGIKELPEPDLTASRLMDAADYVPLKLRCHIDAKD
INSSPKLTLENGTEFFQALIDITHPRRLRYGSGKELAENVRIHRNFTSHRSPPIIYKS
LDNNOYIKYTDVFIIONICGYRKDSTNNIELNRELEVMIYKARETSKRYLFVRE
DADSFRIEKKFEKFNQACFVDRSLQORNSGDHLPIVQELDSMKSIIEFIQHPIORA
KHAAPVIEDLPGSYCILAVALFELFNEIIMWSRYFEKKHVPKIIIEYVAKIPTLLFK
SELBPPIFKTALHRLASTFEVLSFSFHPANPYKEVRNAKADHGFTLQNLKNELTY
LGLETEPDIITDYAEIAYNEVMRYKKEMLLRTCDLDAIEHCQLLCIFNSPFLQFL
HNQACHRLVGLHCDLCIKERKESQEAAMDLESSSTKSPIDCKKCSOTLQTIEDY
KKELTSQNKMKRKVSELEKKLDVENVQNEQVEKEMEDDLRGKQEEINILRNDSL
KVHVDLMEIADLKQHSKLEKHEVEGLKEENLKQERNYNLTQSNLEANSMSREGLTMC
GFSPQYTEELQERQKQEMLTARSEETSVQAEPAEGEEMLDVTQIIEAKEEKIRL
KAHDANERIIQQLDKLANFS"
7825.c.9790
/gene="C16A11.4"
join(7825. .8408,8461. .9120,9306. .9678,9728. .9790)
/gene="C16A11.4"
/note="coded for by C. elegans cDNA yk334c12.3; coded for
by C. elegans cDNA yk298a8.3; coded for by C. elegans cDNA
yk103h8.3; coded for by C. elegans cDNA yk151d2.3; coded
for by C. elegans cDNA yk151d2.5; coded for by C. elegans
cDNA yk103h8.5; coded for by C. elegans cDNA yk298a8.5;
coded for by C. elegans cDNA cm17b3"
/codon_start=1
/protein_id="AAC26260.1"
/db_xref="PID:g3319385"
/db_xref="GI:3319385"
/translation="MDSSEANOFEPPIRSTRVNNLSPHQKEVRQCGTNNNTD
NRLIKVECEKQWHAICMSVPLDIIMTYKCEMCEPRLMVSRSQARAYQGEPEPRE
VRKKGRNOSVNLFELEFLESEASISRLATLTFEDYKEHINYPQSVRIKQLES
NLLETISMPNYADRKQMLFAMGLRYKGEFLERLERNAAVEVDGYHRITFYKSKNL
EFRGKHRVGRFPLVNRKSAESTSQOHNGSEDDDDIFIDLVVPTGNSRWNOFPVK
QEIPDEQSGASTLPNAEIEPNLQYSLSDMTTGSASTWNMAATPPEMGRGIMESM
NGLHALOSASYPYQEMOELCVQHTMMSMHEIITKQSDFYREMEKTRFGWKINFTS
RFSYFHTFFEGYIAEIANRSFYAIAGAPPAPATTSIRAFILHFLKGYVDYLESEQLV
CSRITQOHLTECRADKLISMELTGLAVATTIAEAHLSDVSAIINGASAMEFYKFI
EIIYASNQSNASFGCLCTDLKQIKQIALVNMDDKIPFKELSRAPRAISLVSQ"
10319. .15091
/gene="C16A11.3"
join(10319. .10881,10929. .11022,11160. .11339,11392. .11442,
12018. .12297,12344. .12408,12962. .13039,13096. .13374,
13959. .14338,14659. .14743,14966. .15091)
/gene="C16A11.3"
/note="similar to protein kinases (Pfam: pkinase.hmm,
score: 168.47)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC26259.1"
/db_xref="PID:g3319384"
/db_xref="GI:3319384"
/translation="MEHDSSEDSDFEPPRAPVSSIKIECCGCTNNTNRLIKVCYCA
QWQHAVCMGVPTDCVTAYKCEDEPRLLIVTIAEAESFGYKPEPTPAKIERKTT
KVLVIOREFLDELSDKSENKSPLEPLTKFFQDYKTKSCNSSLATAKKKLLANLEHIV
MSAKYNDNRKAEMLFAIGLVEDEFLELRIGRATVEVDQORITFYKSKTSTSSPEY
SFVLAATTAIAEAHLSDASVIDGSAIEFFKYIENYAKKALLSLCDDLQKIQGLEA
MDKKIYNYLSRVQATISLRKYQKLTGCGEMSKNGNLMAKRFKNOQFFSKMEFPL
KKTFFEQMKRTRAPKEONPSEIKIRKSETDNENAVAGNENLPKNPILITSSGQDLKI
IAKLGEGVGAAYKVVNDKSNKYAVKLFKEHSEVGEQSGMAKAVTAIANSANL
IRMEDSMRINESVPNGFLKIAISFELKGRSVFEVMDSTKGMPAGSEISFEINAIRM
GKQLEAMEKLESVKIVHLDKPENICSSSCTFKTVGNVCFIQSPISFHVAVYDVG
NARTIRQDKPKYALVQTONYRAPEIFLGLPFSVSKSDIWSFGCILSELYTGDLLEYGN
SKSDTEELQELMOMIVQOPSCAMLRKAEKIKSTKRIENGNNVYMKRKSSEFPPK
PLHKQRKPKDLEAIPFLDLECLLIVDPSRRPSLRDISSRTFFK"
complement(15718. .18718)
/gene="C16A11.6"
complement(join(15718. .16143,16238. .16444,16498. .16633,
16683. .16762,17361. .18038,18257. .18392,18438. .18517,

18581. .18718))
/gene="C16A11.6"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC26262.1"
/db_xref="PID:g3319387"
/db_xref="GI:3319387"
/translation="MRNAEKMHADITVNLKGHTDILLHGSRRDVFNRSFLQFLFRI
LTI SANMNSNPITYPSMOCIIKHLANKRFQVLAQCPVLKIEKSQPLIRELSFKRR
TISVNGVEYKLGAKNKFEPHTLAIKOLKGRGPQINYYWLDPPDGRILEQPSFKF
KVRNMDWSSDNPLIDRFIDLKSMPLKIVLSDRYSNLVDVIDRHEILRTSEGLMPH
IVPRELFHGIGTLSTELIVANIEGVEFDFLNLIDALLQPDGRNGADRSFTEDADLI
ANLEKAFTLAEGKDRAISEYEGEDECFFRIRAEYTKCDEFPNEIGFENDYMIHA
DAKTEENKFNISAKMNSNPITYPSMOCIIKHLANKRFQVLAQCSSLVAYEKSQPLVI
KLILFNRTISFNGVKKYLGAKNKFEPHTLAIKQKGRGPQINVOWLDFPDGRILE
LPSEFKKQVQHFMTRSSDNPLIERFIDPKSLPKNCSTPSEGTLPALIRAGLDNDA
EYFDFLNLIDALLQDREYTGRCFITEDVDLVSELEFAFTLLAEKDPATYEEDEN
FDFEFEDGYTRCDEFPNEIAFWENDDMVISADVYKEKDDFQWLWLRREEKTLDDLFIN
YF"
18981. .25463
/gene="C16A11.2"
join(18981. .19013,19054. .19152,19203. .19306,20074. .20557,
21287. .21579,22919. .23371,24089. .24284,25173. .25463)
/gene="C16A11.2"
/note="coded for by C. elegans cDNA yk275g3.5"
/codon_start=1
/protein_id="AAC26258.1"
/db_xref="PID:g3319383"
/db_xref="GI:3319383"
/translation="MDLNAAAAEKVPETSSNGRNDIIAMLSAGYHRAVSCINDYDK
IVGGIWAIMRCDVDVAVDVLLEYESEDQNTKLRTEQSEKLIIEALKAVQCPHEIATHQ
ISGLDFLALKNVLIOWLLRHLVLAQESKGNKRFHFTWFSMHHEHRDEALKDERHHLRG
VTTTRYMKRFDNSAMFNLSMDAKCTLAIEYVYVSRGTDRLVGFTEDEETDNLKWLKST
ILSEKLEKIRFENFNLNRKFEENASTKRVPTTVRKMDAEAVETFEKNDTESIKTST
EGMCKEVEDELNIQORLEALESVIDEAELEQOVKDTHKQLSDRYNTIKNRVNSNDOMF
TFYNSNSDIDQLAETVKEFEVTOARADELRACACDVRMAONHENILRONDKNGKMQ
KYCVKQREIVDMAERMKMYVTCDRVVALQFRLDRLKTLISHYKRHMEKLELSE
LTOEAKTNVIDHNTVIDILTFSAKIKGFENDEVEKSLVPEASQERDAFVSYMNDRVT
QUGEYHWKAKLTQDKTLAEKDTLALRTRIRQKEREVSNTSDEMRLKHFEKVLKNC
KLFPYRLGNLIEFEFEINSKNNSKFSKSPKISKKKFQMLKIYESLKEVDSLLAQR
EISPAFOYALDSREAEDEDEEAEEES"
33481. .36049
/gene="C16A11.1"
join(33481. .33632,33795. .34271,34988. .35116,35232. .35274,
35582. .36049)
/gene="C16A11.1"
/note="similar to the G-alpha family of guanine
nucleotide-binding proteins (Pfam: G-alpha.hmm, score:
410.24)"
...
Note: remainder of annotations omitted.
Query Match 4.8%; Score 48; DB 21; Length 47147;
Best Local Similarity 79.3%; Pred. No. 6,66e-09;
Matches 65; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
Db 24724 GAAATTAGTGTAAATATAAAAAATTACTTGATTATCTGAAAATAAAGTTATTCA 24783
Cp 2766 GAAATTTAGTGTAAATAATAATACCTGAGCTTTTCAGTAATAATAATATTTC 2707
Db 24784 GCTGAAAAAATTCGAGAAAGTGC 24805
Cp 2706 CTTCAAATATTTTGGGAAAGTGC 2685
RESULT 6
LOCUS CELF22D3 35713 bp DNA INV 15-JUN-1995
DEFINITION Caenorhabditis elegans cosmid F22D3.
ACCESSION U28993
NID 9861377
VERSION U28993.1 GI:861377
KEYWORDS
SOURCE Caenorhabditis elegans strain-Bristol N2.

DB	Accession	Sequence	Length	Source	Reference	Patent	Date
Db	240992	GTGAAGTCAGTAAATCGGAAAAATTTTGACCCCTCTGGAATTTTATCTGATTTCTGG	166494	7218 bp	DNA	PAT	23-DEC-1997
CP	2020	GTGAAAGTCAGTAAATCGGAAAAATTTTGACCCCTCTGGAATTTTATCTGATTTCTGG	166494	7218 bp	DNA	PAT	23-DEC-1997
Db	241052	CGAATTTTGGGGTTTTTTTTTTGTTAAATTTGTTGAATCAGTCTCTTTAATTTTCGCA	166494	7218 bp	DNA	PAT	23-DEC-1997
CP	1960	CGAATTTTGGGGTTTTTTTTTTGTTAAATTTGTTGAATCAGTCTCTTTAATTTTCGCA	166494	7218 bp	DNA	PAT	23-DEC-1997
Db	241112	AACATAAAACACAAATTTTCAATGAGAAATCGTTATTTTAAATTTGAAAACTGATTTTAA	166494	7218 bp	DNA	PAT	23-DEC-1997
CP	1900	AACATAAAACACAAATTTTCAATGAGAAATCGTTATTTTAAATTTGAAAACTGATTTTAA	166494	7218 bp	DNA	PAT	23-DEC-1997
Db	241172	AATGTTTGTGTGTTTTTTGTTCTGTTTGTACGCCCTGGAA	166494	7218 bp	DNA	PAT	23-DEC-1997
CP	1840	AATGTTTGTGTGTTTTTTGTTCTGTTTGTACGCCCTGGAA	166494	7218 bp	DNA	PAT	23-DEC-1997
LOCUS	4	166494	7218 bp	DNA	PAT	23-DEC-1997	
DEFINITION		Sequence 14 from patent US 5670367.					
ACCESSION		166494					
NID		92724471					
VERSION		166494.1					
KEYWORDS		GI:2724471					
SOURCE		Unknown.					
ORGANISM		Unknown.					
REFERENCE		Unclassified.					
AUTHORS		1 (bases 1 to 7218)					
TITLE		Dorner, F., Scheiflinger, F. and Falkner, F. Gunter.					
JOURNAL		Recombinant fowlpox virus					
FEATURES		Patent: US 5670367-A 14 23-SEP-1997;					
		Location/Qualifiers					
		1..7218					
BASE COUNT		1944 a 1491 c 1486 g 1929 t 368 others					
ORIGIN		/organism="unknown"					

	Query Match	7.6%;	Score 76;	DB 25;	Length 7218;	
	Best Local Similarity	0.9%;	Pred. No. 1.87e-23;			
	Matches	3;	Conservative	204;	Mismatches 131;	Indels 0; Gaps 0;
Db	1063 CGATYYY	1122				
Cp	2502 CGAAGCCACATCTCCCCGTCGTGCAGCTGCTTACGATCTCCCGTCTCTCGCGAA	2443				
Db	1123 YYY	1182				
Cp	2442 CCGTTCCACATGACTGCAAAATAATATTCTTTATATTCAATTTTTTTAAAAATTATAT	2383				
Db	1183 YYY	1242				
Cp	2382 TCGATTAAAAAACAATCACAATTAATCATATCTCCATTAATCACTATTCAGACTTGTTGGCA	2323				
Db	1243 YYY	1302				
Cp	2322 TGAGAAGCTCGAGAATTTCATCGACCTTTAGATGACTAGAGAACATCATTAATGTTCTCT	2263				
Db	1303 YYY	1362				
Cp	2262 CTAGCAAGCTCTCTTCTATCTTGACGCATCACTTCGGCTGATGGTTGACCTCTACTGA	2203				
Db	1363 YYY	1400				
Cp	2202 CCCGTAGTCAAGTACGATTTGATTTTCGATTTAGTCT	2165				

RESULT	5		INV	15-JUL-1998
LOCUS	CEL216A11	47147 bp	DNA	
DEFINITION	Caenorhabditis elegans	cosmid	C16A11.	
ACCESSION	AF077536			
NID	93319380			
VERSION	AF077536.1	GI:3319380		

KEYWORDS	
SOURCE	
ORGANISM	
	Caenorhabditis elegans.
	Caenorhabditis elegans
	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
	Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE	
AUTHORS	1 (bases 1 to 47147) Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Shownkeen, R., Smaildon, N., Smith, A., Sonnenhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J. and Wohlman, P.
TITLE	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL	Nature 368 (6466), 32-38 (1994)
MEDLINE	94150718
REFERENCE	2 (bases 1 to 47147)
AUTHORS	Johnson, D. and Biewald, T.
TITLE	The sequence of C. elegans cosmid C16A11
JOURNAL	Unpublished (1998)
REFERENCE	3 (bases 1 to 47147)
AUTHORS	Waterston, R.
TITLE	Direct Submission
JOURNAL	Submitted (13-JUL-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, MO 63108, USA
COMMENT	Submitted by:

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is F31D5, 5100 bp overlap; 3' cosmid is C01F1, 200 bp overlap. Actual start of this cosmid is at base position 5097 of CELC16a11; actual end is at 1945 of CELC01F1

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES
Source

gene 4082. .7021
/gene="C16A11.5"
join(4082. .4186,4248. .4356,4420. .4487,4572. .4676,
CDS 4735. .4899,4978. .5038,5443. .5666,5725. .6048,6095. .6447,

CP 2560 TCAAGAACTTCAGCAAGTCTTCTGTCGCCGTAGACGGAAGCATCATATAAACGGCTCG 2501
Db 10855 AACGCCACATCTCCCGCTGCTGACACTGCTTTACGATCTCCCGTCTCTCTCGGAACC 10914
CP 2500 AACGCCACATCTCCCGCTGCTGACACTGCTTTACGATCTCCCGTCTCTCTCGGAACC 2441
Db 10915 GTTCCACATGACTGCAAAAATATTTCTTATATATCAATTTTAAATTTATTTATTC 10974
CP 2440 GTTCCACATGACTGCAAAAATATTTCTTATATATCAATTTTAAATTTATTTATTC 2381
Db 10975 GATTAAAACTCACATTAATCATATCTCCATATATCACTATTCAACACTGTTTGGCATG 11034
CP 2380 GATTAAAACTCACATTAATCATATCTCCATATATCACTATTCAACACTGTTTGGCATG 2321
Db 11035 AGAAGCTCGAGAAATTCATGCACTTTAGATGACTAGAGAACATCATATATGTTCTCTCT 11094
CP 2320 AGAAGCTCGAGAAATTCATGCACTTTAGATGACTAGAGAACATCATATATGTTCTCTCT 2261
Db 11095 AGCAAGCTCCTCTATCTTGACGCAATCTTCGGCTGATGGTTCACCTCTCTACTGACC 11154
CP 2260 AGCAAGCTCCTCTATCTTGACGCAATCTTCGGCTGATGGTTCACCTCTCTACTGACC 2201
Db 11155 CGTAGTCAGAGTACGATTTGATTTTCGATTTAGTCTGAAAATGATGAATTTATTTGAAA 11214
CP 2200 CGTAGTCAGAGTACGATTTGATTTTCGATTTAGTCTGAAAATGATGAATTTATTTGAAA 2141
Db 11215 AATAGTGTCTTTTAAACATTTGCTCAAAATTCAGCTCATTTGAGTAATTTTATTC 11274
CP 2140 AATAGTGTCTTTTAAACATTTGCTCAAAATTCAGCTCATTTGAGTAATTTTATTC 2081
Db 11275 GATTTTCTCTAGTTTTCGTCAAAAACCTCCACACACTGCACTGAACTAGCAAAAAAAG 11334
CP 2080 GATTTTCTCTAGTTTTCGTCAAAAACCTCCACACACTGCACTGAACTAGCAAAAAAAG 2021
Db 11335 GTGAAGTCAGTAAATCGGAAATTTTGACCCCTCTGAAATTTTATCTTGAATTTCTGG 11394
CP 2020 GTGAAGTCAGTAAATCGGAAATTTTGACCCCTCTGAAATTTTATCTTGAATTTCTGG 1961
Db 11395 CGAATTTTGGGGTTTCTTGTAAATTTGTTGATTCAGTCTCTTTATTTTTCGCA 11454
CP 1960 CGAATTTTGGGGTTTCTTGTAAATTTGTTGATTCAGTCTCTTTATTTTTCGCA 1901
Db 11455 AACATAAACACATTTTCAATGAGAAATCGTATTTTAAATTTGAAAACGATTTTAA 11514
CP 1900 AACATAAACACATTTTCAATGAGAAATCGTATTTTAAATTTGAAAACGATTTTAA 1841
Db 11515 AATGTTTGTGTTTGTCTGTTGTCAGCTTGAA 11554
CP 1840 AATGTTTGTGTTTGTCTGTTGTCAGCTTGAA 1801

RESULT 3
LOCUS CEY67H2 314495 bp DNA HTG 04-MAR-1999
DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
clone Y67H2, WORKING DRAFT SEQUENCE.

ACCESSION AL022475
NID 94469034
VERSION AL022475.3 GI:4469034
KEYWORDS HTG; HTGS PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.

REFERENCE 1 (bases 1 to 314495)
AUTHORS McMurray, A.
TITLE Direct Submission
JOURNAL Submitted (04-MAR-1999) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridgeshire CB10 1RO, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or twenematode.wustl.edu

COMMENT On Mar 22, 1999 this sequence version replaced gi:4468145.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress

and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source Location/Qualifiers
1. .314495
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="Y67H2"

BASE COUNT 96542 a 54551 c 52760 g 97027 t 13615 others
ORIGIN

Query Match 99.1%; Score 991; DB 19; Length 314495;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 998; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 240213 TTCGGCAAAATCTACAAATATGCGCTTTGCGCCACGAAATATAGTGTAAAAATAATAA 240272
CP 2800 TTCGGCAAAATCTACAAATATGCGCTTTGCGCCACGAAATATAGTGTAAAAATAATAA 2741
Db 240273 TCACCTGAGCTTTCAGTAATATAATATATTCTACTTCAAAATATTGGGAAAGTGCA 240332
CP 2740 TCACCTGAGCTTTCAGTAATATAATATATTCTACTTCAAAATATTGGGAAAGTGCA 2681
Db 240333 TTCCGATTTTCGGAGGAAAAAACCGCTGCCGGTGGGTGCAAAATTTACGATCCGCTCTT 240392
CP 2680 TTCCGATTTTCGGAGGAAAAAACCGCTGCCGGTGGGTGCAAAATTTACGATCCGCTCTT 2621
Db 240393 TTT-CCGTGAGACCCCTTGCTTTTGCGCGCAACTTTAAACCTACGATCTGGCGAGAGT 240451
CP 2620 TTTCCGTGAGACCCCTTGCTTTTGCGCGCAACTTTAAACCTACGATCTGGCGAGAGT 2561
Db 240452 TCAAGAACTTCAGCAAGTCTCTGCTGCTCCCGTAGAGCGAAGACATCATATAAACGCTCG 240511
CP 2560 TCAAGAACTTCAGCAAGTCTCTGCTGCTCCCGTAGAGCGAAGACATCATATAAACGCTCG 2501
Db 240512 AACGCCACATCTCCCGCTGCTGCACTGCTTTCAGCATCTCCCGTCTCTTCGCGGAACC 240571
CP 2500 AACGCCACATCTCCCGCTGCTGCACTGCTTTCAGCATCTCCCGTCTCTTCGCGGAACC 2441
Db 240572 GTTCCACATGACTGCAAAATAATTTCTTATATATCAATTTTAAATTTATTTATTC 240631
CP 2440 GTTCCACATGACTGCAAAATAATTTCTTATATATCAATTTTAAATTTATTTATTC 2381
Db 240632 GATTAAAACTCACATTAATCATATCTCCATATATCACTATTCAACACTGTTTGGCATG 240691
CP 2380 GATTAAAACTCACATTAATCATATCTCCATATATCACTATTCAACACTGTTTGGCATG 2321
Db 240692 AGAAGCTTCGAGAAATTCATGCACTTTAGATGACTAAAGAACATCATATGTTCTCTCT 240751
CP 2320 AGAAGCTTCGAGAAATTCATGCACTTTAGATGACTAAAGAACATCATATGTTCTCTCT 2261
Db 240752 AGCAAGCTCCTCTATCTTGACGCAATCTTCGGCTGATGGTTCACCTCTCTACTGACC 240811
CP 2260 AGCAAGCTCCTCTATCTTGACGCAATCTTCGGCTGATGGTTCACCTCTCTACTGACC 2201
Db 240812 CGTAGTCAGAGTACGATTTGATTTTCGATTTAGTCTGAAAATGATGAATTTATTTGAAA 240871
CP 2200 CGTAGTCAGAGTACGATTTGATTTTCGATTTAGTCTGAAAATGATGAATTTATTTGAAA 2141
Db 240872 AATAGTGTCTTTTAAACATTTGCTCAAAATTCAGCTCGATTTGAGTAATTTTATTC 240931
CP 2140 AATAGTGTCTTTTAAACATTTGCTCAAAATTCAGCTCGATTTGAGTAATTTTATTC 2081
Db 240932 GATTTTCTCTAGTTTTCGTCAAAAACCTCCACAACTGCACTGAACTAGCAAAAAAAG 240991
CP 2080 GATTTTCTCTAGTTTTCGTCAAAAACCTCCACAACTGCACTGAACTAGCAAAAAAAG 2021

REFERENCE 2 (bases 1 to 39908)
AUTHORS Wilson,R., Alnsough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Showkeen,R., Smaiden,N., Smith,A., Sonhammer,E., Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkison-Sproat,J. and Wohlman,P.
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
COMMENT Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.
For a graphical representation of this sequence and its analysis see:-
http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence&object=C48D1
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
This sequence is the entire insert of clone C48D1. The true right end of clone F58D2 is at 18510 in this sequence. The start of this sequence (1..100) overlaps with the end of sequence 281093.
The end of this sequence (33020..39908) overlaps with the start of sequence 282274.

FEATURES
source Location/Qualifiers
1. 39908
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="C48D1"
complement(94..4439)
/gene="C48D1.1"
complement(join(94..1020,1345..1951,2283..2580,2629..2765,2882..3147,3615..3739,4089..4331,4382..4439))
/gene="C48D1.1"
/note="similar to C.elegans ZYG-11 like protein; cDNA EST EMBL:Z14909 comes from this gene; cDNA EST EMBL:T01768 comes from this gene"
/codon_start=1
/protein_id="CAB02846.1"
/db_xref="PID:e1344741"
/db_xref="PID:g3875048"
/db_xref="GI:3875048"
/db_xref="SPTREMBL:O02227"
/translation="MTKPLHDIACQIAEFVQNGSYDNREFELDSKSSNIYSYLMNRDKPFLKTHKLVNTKVTLPQTDENTAROLKOFSLKELIGNFSKSKSWCIFRKE TRHVDIVRIEAVLENEETRNLYHIEKDQKIMKRTAKIGNMLPSLTHLMSDVHI NEDDFYQLGNTFNSRLSDISGTNINIAISRLKOLEVALRDESIATYTELKDLFN LKHLRLVDVSOQRENLHSRNINIVEQTECKMELPELRMDCTRTSLSGDILECLLKT LSIPNIIIVTEAEATSOYNDVRVNTAFYHKSVALVYCLTARKEFDSGYCMKAIRISI SEDNLOINIPSCIKITIVKSMQTFPSISVTQNGKLCILTFGLYSNOLCSNGVTLIVE ALMSFVLYLKRMISEAVYKVNISYNNAVEKLVNKLKGLBDEKISKATMSFLENRVS KALRTASSVLCOCIRHMSDELKSEKTEVNRLVDLEWSVGQKNETKSEFEI TIKMTEKSSKACTNMIAQGISISLSCEKVESNDITTKCLELLITNLSALNDSLQEL YTSNNDKIGONAYSTEK"
complement(6258..11120)
/gene="C48D1.2"
complement(join(6258..6349,6876..7065,7420..7547,7602..7815,8728..9059,10248..10501,10780..10925,10989..11120))

gene
CDS
gene
CDS

/gene="C48D1.2"
/note="similar to Caspase recruitment domain, ICE-like protease (caspase) p20 domain.; ICE-like protease (caspase) p10 domain.; cDNA EST EMBL:D75533 comes from this gene; cDNA EST yk198f10.5 comes from this gene; cDNA EST yk491a5.5 comes from this gene"
/codon_start=1
/protein_id="CAB02848.1"
/db_xref="PID:e1344743"
/db_xref="PID:g3875050"
/db_xref="GI:3875050"
/db_xref="SPTREMBL:O02229"
/translation="MRDRSLERINIMFSSHLKVDLELVIAKOVLSNDGMDMIN SGTVREKRREIVAVQGRGDVAFDAFYDALRSTGHEGLAEVLEPLARSDSNAVEFE CPMSPAHSRRSRAISPAGYTSPTRVHRDSVSSVSTSYQDIYSRARSRSRALHSS DRHNYSSPVNAFPSPQSSASSFTGSSGYSSSRNRSFSKASGPTQYIFHEEDMNF VDAPTISRVFDEKTMTRNFSSPRGMLITNEHEFQMPTRNGTKADKDLTNLFRCMG YVYICKDLNLTGRGMLTIRDEFAKHESHGSAIIVLSHGEENVYIIGVDIPSTHEIY DLNANAPRLANKPKIVFVQACRGERDNGFPVLDSDVGPAPFLRGWDRDGPLFN FLGCVRPQVQVWRKRKPSQADILIAVATYQVSVSRNSARGSWFIQAVCEVFSTHAKD MDVVELITEVNKKVAYDIPPAQKVLVLLAGSTKLCRLKFTKRLDLPN"
complement(16940..21616)
/gene="C48D1.3"
complement(join(16940..17126,18419..18597,18985..19442,19553..19879,19925..19961,20333..20481,20789..20987,21037..21152,21373..21616))
/gene="C48D1.3"
/note="Similarity to Salmonella sodium/proline symporter (SW:PUTP_SALTY); cDNA EST EMBL:C13840 comes from this gene; cDNA EST EMBL:C11562 comes from this gene"
/codon_start=1
/protein_id="CAB02847.1"
/db_xref="PID:e1344742"
/db_xref="PID:g3875049"
/db_xref="GI:3875049"
/db_xref="SPTREMBL:O02228"
/translation="MADLGIYAIIVEFYVLIVGIMAGRSKSKSELEAGATEE VMLAGRNIGTLVGIFTMTATWVGAIYINGTAELVNGLLGCQAPVGVATSLVMGGL FAKMREGYITMLDPQFQWNLLELIGRTFDNFRKRLKQIOTIIEIDFQHKYG ORIGGLMYVPALIGETFWTALISALGATLSVILGIDMNASVTLACIAVFTFTGGY YAVAYTDVQVLCIFVGLLIGLVQNRPNRFKETSUIDCMLLVFGGIPWQVYFOR VLSSKTAHGAQTLSEFVAGVGCILMAIPALIGAIARNTDWRMTDYSFWMNGTKVESIP PDKRMNVVPLVFQYLTFRWVAFGLGAVSAVMSADSSVLSAASFANIMKLTIRP HASEKEVITIVRIALICVIGIMATIMALTIOSIYGLWYLCADLVYVILFPOLICVYMP RSNYGLSLAGYAVGLVRLIGCEPLVSLPAFFHYPMYTDGQYFERTTAMSSMATI YVVISQSEKLEKSGRLSPEDWGVNIPIDHVPPLPSVDFSVASSETLNMYECDDM QEPOLQTEHRLQYIRTSRLMKIHYIIRTKVITIPQAIKRDLELVRVHLISQIIEF CIFSFR"

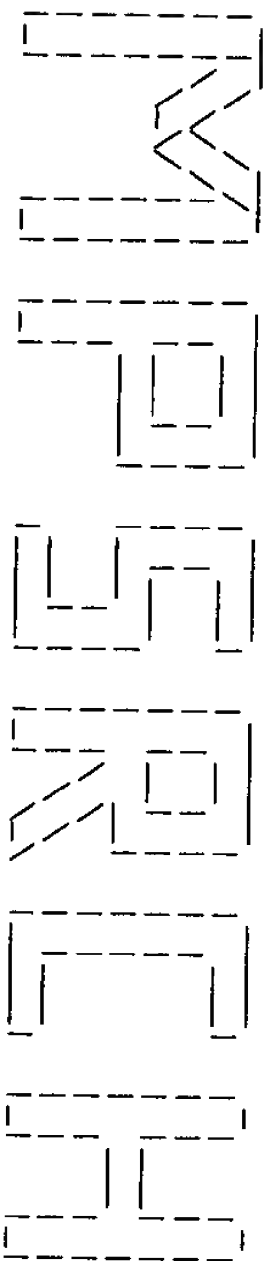
BASE COUNT 13182 a 6945 c 7146 g 12635 t
ORIGIN
Query Match 99.3%; Score 993; DB 21; Length 39908;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 999; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Db 10556 TTGGGCAATCTACAAATGCGCTTTTGGCGCACGAAATTAAGTGTATAAAATAAATAA 10615
Cp 2800 TTGGGCAATCTACAAATGCGCTTTTGGCGCACGAAATTAAGTGTATAAAATAAATAA 2741
Db 10616 TCACCTGAGCTTTCAGTAAATAAATATATTTCACCTTCAATATTTGGGAAGTCATAA 10675
Cp 2740 TCACCTGAGCTTTCAGTAAATAAATATATTTCACCTTCAATATTTGGGAAGTCATAA 2681
Db 10676 TTTCCGATTTTCGAGGAGGAAAAACCGTGCCTGGGTTGCAAAATTAAGATCCGCTCTT 10735
Cp 2680 TTTCCGATTTTCGAGGAGGAAAAACCGTGCCTGGGTTGCAAAATTAAGATCCGCTCTT 2621
Db 10736 TTT-CCGTGAGACCTTGTCTTTTGGCGGAACTTTAAAAAAGCTAGATCTGGCGAGAGGT 10794
Cp 2620 TTTTCCGTGAGACCTTGTCTTTTGGCGGAACTTTAAAAAAGCTAGATCTGGCGAGAGGT 2561
Db 10795 TCAAGAACTTCAGCAAGTCTCTGCTGCCCTAGAGCGAAGACATCATAAAAACCGCTCG 10854
|||||

|||||


```
/protein_id="AAA27982.1"
/db_xref="PID:g456417"
/db_xref="GI:456417"
/translation="MMRODRSLLEERNIMFSSHLKVEILEVILAKOVLSNDGMI
NSCTVREKREIVKAVQRRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDNAVEF
ECMPSPASHRSRSLSPAGTSPTRVHRDSVSSVSTSYODIYSRARSRSRSLHS
SDRHNSSPPVNAFPPSQSSANSSFTGCSLSGYSRRNSFSKASGPTQYIFHEEDM
FVDAPITSRVEDEKMTNRNSSFPGMCLINNEHEQMPTRNGTKADNLINLFRM
GTVICKDNLTRGMLLTIRDFAKHSHGSDAILVILSHGEENVIGVDDIPISTHEI
YDLNANANAPRLANKPKIIVEQACRGERRDNGEPYLDSDGVPAPFLRGMDNDGPLF
NELGCVRPQOVWRKKPSQADILIRYATTQYVSWMNSARSGSWFIQAVCEVSTHAK
DMDVVELITEVNKKVACGFGTSGSNILKQPEMTSRLKKFYFWPEARNASAV"
intron
2367. .2429
/gene="ced-3"
/number=1
exon
2430. .2575
/gene="ced-3"
/number=2
intron
2576. .2853
/gene="ced-3"
/number=2
exon
2854. .3107
/gene="ced-3"
/number=3
intron
3108. .4302
/gene="ced-3"
/number=3
exon
4303. .4634
/gene="ced-3"
/number=4
intron
4635. .5546
/gene="ced-3"
/number=4
exon
5547. .5760
/gene="ced-3"
/number=5
intron
5761. .5814
/gene="ced-3"
/number=5
exon
5815. .5942
/gene="ced-3"
/number=6
intron
5943. .6297
/gene="ced-3"
/number=6
exon
6298. .6537
/gene="ced-3"
/number=7
intron
6538. .7012
/gene="ced-3"
/number=7
exon
7013. .7652
/gene="ced-3"
/number=8
BASE COUNT 2429 a 1455 c 1271 g 2498 t
ORIGIN
Query Match 100.0%; Score 1000; DB 21; Length 7653;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Db 1981 TTCAGAGGGTCAAAATTTTCCGATTTTACTGACTTTCACCTTTTTCGTAGTCACT 2040
QY 1981 TTCAGAGGGTCAAAATTTTCCGATTTTACTGACTTTCACCTTTTTCGTAGTCACT 2040
Db 2041 GCAGTGTGGAGTTTTCAGCAAACTAGCAAAAAATCGATAAAAATTTACTCAAAATCG 2100
QY 2041 GCAGTGTGGAGTTTTCAGCAAACTAGCAAAAAATCGATAAAAATTTACTCAAAATCG 2100
Db 2101 AGCTGAATTTTGAGACAAATGTTTAAAAAAAACACTATTTTCCAAATATTCTACTCAT 2160
QY 2101 AGCTGAATTTTGAGACAAATGTTTAAAAAAAACACTATTTTCCAAATATTCTACTCAT 2160
Db 2161 TTCAGACTAAATCGAAAAATCAAAATCGTACTCTGACTACGGTCACTAGAGAGGTCACCC 2220
QY 2161 TTCAGACTAAATCGAAAAATCAAAATCGTACTCTGACTACGGTCACTAGAGAGGTCACCC 2220
Db 2221 ATCAGCCGAGATGATGCGTCAAGATAGAGAGAGCTTCTAGAGAGGAACTATGATGT 2280
QY 2221 ATCAGCCGAGATGATGCGTCAAGATAGAGAGAGCTTCTAGAGAGGAACTATGATGT 2280
Db 2281 TCTCTAGTCACTAAAGTGCATGAATTTCTGAAAGTCTCATCGCAAAACAAGTGTGA 2340
QY 2281 TCTCTAGTCACTAAAGTGCATGAATTTCTGAAAGTCTCATCGCAAAACAAGTGTGA 2340
Db 2341 ATAGTATATGAGATATGATTAATGTGAGTTTAAATCGAATATTAATTTAAAAAAA 2400
QY 2341 ATAGTATATGAGATATGATTAATGTGAGTTTAAATCGAATATTAATTTAAAAAAA 2400
Db 2401 AATTGATATATTAAGAATATTTTTCAGTCACTATGTGAACGGTTCCGAGAGAGACGGG 2460
QY 2401 AATTGATATATTAAGAATATTTTTCAGTCACTATGTGAACGGTTCCGAGAGAGACGGG 2460
Db 2461 AGATCGTGAAGCAGTGCACACGACGGGAGATGTGGCGTTTCAGCGGTTTATGATGCTC 2520
QY 2461 AGATCGTGAAGCAGTGCACACGACGGGAGATGTGGCGTTTCAGCGGTTTATGATGCTC 2520
Db 2521 TTCGCTCTACGGGACGACGAGGACTTGTGAAGTCTTGAACCTCTCCGACAGATGCTAGG 2580
QY 2521 TTCGCTCTACGGGACGACGAGGACTTGTGAAGTCTTGAACCTCTCCGACAGATGCTAGG 2580
Db 2581 TTTTAAAGTTCGGCGCAAAAGCAAGGGTCTCACGAAAAAAGAGCGGATCGTAATTTT 2640
QY 2581 TTTTAAAGTTCGGCGCAAAAGCAAGGGTCTCACGAAAAAAGAGCGGATCGTAATTTT 2640
Db 2641 GCAACCCACCGGACGAGTTTTCCTCCGAAAAATCGGAAATATGACCTTTCCCAAAATAT 2700
QY 2641 GCAACCCACCGGACGAGTTTTCCTCCGAAAAATCGGAAATATGACCTTTCCCAAAATAT 2700
Db 2701 TTGAAGTGAATATATTTTACTGAAAGCTGAGTATTTTAACTTTTAACTA 2760
QY 2701 TTGAAGTGAATATATTTTACTGAAAGCTGAGTATTTTAACTTTTAACTA 2760
Db 2761 ATTTTCGTGGCGCAAAAGGCAATTTTGTAGATTGCGGAA 2800
QY 2761 ATTTTCGTGGCGCAAAAGGCAATTTTGTAGATTGCGGAA 2800
```

```
RESULT 2
LOCUS CEC48D1 39908 bp DNA INV 23-NOV-1998
DEFINITION Caenorhabditis elegans cosmid C48D1, complete sequence.
ACCESSION Z81049
NID 91627677
VERSION Z81049.1 GI:1627677
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 39908)
AUTHORS Burton, J.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rw@nematode.wustl.edu
```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_nm n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Fri Aug 6 03:10:58 1999; MasPar time 1828.29 Seconds

Tabular output not generated. 1516.122 Million cell updates/sec

Title: >US-08-287-669-18
Description: (1801-2800) from US08287669.seq (3 of 10)

Perfect Score: 1000
N.A. Sequence: 1801 TTCACGGCTGACAAACAGAA.....CATTTGTAGATTGCGCGAA 2800
Comp: AAGGTCGACACTGTTGTCTT.....GTAAACATCTTAAACGCGCTT

Scoring table: TABLE default

Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

emb158

Database:

genbank111

17:gb_ba1 18:gb_ba2 19:gb_htg1 20:gb_htg2 21:gb_in1
22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
28:gb_pl2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_vi

Statistics: Mean 11.320; Variance 7.832; scale 1.445

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1000	100.0	7653	21	CELCED3A	Caenorhabditis elegans	0.00e+00
2	993	99.3	39908	21	CEC48D1	Caenorhabditis elegans	0.00e+00
3	991	99.1	314495	19	CEY67H2	Caenorhabditis elegans	0.00e+00
4	76	7.6	7218	25	166494	Sequence 14 from paten	1.87e-23
5	48	4.8	47147	21	CELC16A11	Caenorhabditis elegans	6.66e-09
6	45	4.5	35713	21	CELF22D3	Caenorhabditis elegans	1.92e-07
7	44	4.4	10772	21	AF012089	Drosophila melanogaste	5.82e-07
8	44	4.4	41564	21	CEY53F4	Caenorhabditis elegans	5.82e-07
9	44	4.4	110000	19	CEY53F4	Caenorhabditis elegans	5.82e-07
10	43	4.3	10772	21	AF012089	Drosophila melanogaste	1.75e-06
11	43	4.3	33164	21	CEY35C5	Caenorhabditis elegans	1.75e-06
12	43	4.3	31530	19	CEY46G5	Caenorhabditis elegans	1.75e-06
13	42	4.2	26276	21	CELF53E10	Caenorhabditis elegans	5.20e-06

14	42	4.2	145920	19	AC006710	Caenorhabditis elegans	5.20e-06
15	42	4.2	224746	19	CEY56A3	Caenorhabditis elegans	5.20e-06
16	42	4.2	249640	19	CEY79H2	Caenorhabditis elegans	5.20e-06
17	42	4.2	299919	19	AC006712	Caenorhabditis elegans	5.20e-06
18	41	4.1	22881	21	CEZC518	Caenorhabditis elegans	1.54e-05
19	41	4.1	39973	21	CEM106	Caenorhabditis elegans	1.54e-05
20	41	4.1	39973	21	CEM106	Caenorhabditis elegans	1.54e-05
21	41	4.1	74371	31	AC005369	Homo sapiens chromosome	1.54e-05
22	41	4.1	110000	19	CEY11B2	Caenorhabditis elegans	1.54e-05
23	41	4.1	125590	21	CEY49E10	Caenorhabditis elegans	1.54e-05
24	41	4.1	179854	19	CEY19D2	Caenorhabditis elegans	1.54e-05
25	41	4.1	260699	20	CEY19D2	Caenorhabditis elegans	1.54e-05
26	41	4.1	38166	21	CEY43G9	Caenorhabditis elegans	1.54e-05
27	40	4.0	145920	19	AC006893	Caenorhabditis elegans	4.50e-05
28	40	4.0	299919	19	AC006710	Caenorhabditis elegans	4.50e-05
29	40	4.0	299919	19	AC006712	Caenorhabditis elegans	4.50e-05
30	39	3.9	11230	21	CEY102A5A	Caenorhabditis elegans	1.30e-04
31	39	3.9	25600	21	CELT20F5	Caenorhabditis elegans	1.30e-04
32	39	3.9	29249	21	CEC18D11	Caenorhabditis elegans	1.30e-04
33	39	3.9	36355	21	CEM01F1	Caenorhabditis elegans	1.30e-04
34	39	3.9	42614	21	CEB0513	Caenorhabditis elegans	1.30e-04
35	39	3.9	75073	21	CEY48A6B	Caenorhabditis elegans	1.30e-04
36	39	3.9	110000	19	CEY102A5	Caenorhabditis elegans	1.30e-04
37	39	3.9	224746	19	CEY56A3	Caenorhabditis elegans	1.30e-04
38	39	3.9	244513	19	CEY80D3	Caenorhabditis elegans	1.30e-04
39	39	3.9	249640	19	CEY79H2	Caenorhabditis elegans	1.30e-04
40	39	3.9	261002	19	CEY41C4	Caenorhabditis elegans	1.30e-04
41	39	3.9	296699	21	CEY48A6	Caenorhabditis elegans	1.30e-04
42	39	3.9	298406	21	CEY75B8A	Caenorhabditis elegans	1.30e-04
43	39	3.9	316170	19	CEY37A1	Caenorhabditis elegans	1.30e-04
44	39	3.9	336638	19	CEY75B8	Caenorhabditis elegans	1.30e-04
45	38	3.8	244513	19	CEY80D3	Caenorhabditis elegans	3.75e-04

ALIGNMENTS

RESULT	1	LOCUS	7653 bp	DNA	INV	23-FEB-1994
DEFINITION	1	CELCED3A	7653 bp	DNA	INV	23-FEB-1994
ACCESSION	1	CELCED3A	7653 bp	DNA	INV	23-FEB-1994
VERSION	1	L29052	7653 bp	DNA	INV	23-FEB-1994
KEYWORDS	1	L29052	7653 bp	DNA	INV	23-FEB-1994
SOURCE	1	L29052	7653 bp	DNA	INV	23-FEB-1994
ORGANISM	1	L29052	7653 bp	DNA	INV	23-FEB-1994
REFERENCE	1	L29052	7653 bp	DNA	INV	23-FEB-1994
AUTHORS	1	L29052	7653 bp	DNA	INV	23-FEB-1994
TITLE	1	L29052	7653 bp	DNA	INV	23-FEB-1994
JOURNAL	1	L29052	7653 bp	DNA	INV	23-FEB-1994
MEDLINE	1	L29052	7653 bp	DNA	INV	23-FEB-1994
FEATURES	1	L29052	7653 bp	DNA	INV	23-FEB-1994
source	1	L29052	7653 bp	DNA	INV	23-FEB-1994
exon	1	L29052	7653 bp	DNA	INV	23-FEB-1994
gene	1	L29052	7653 bp	DNA	INV	23-FEB-1994
CDS	1	L29052	7653 bp	DNA	INV	23-FEB-1994
Location/Qualifiers	1	L29052	7653 bp	DNA	INV	23-FEB-1994
/organism="Caenorhabditis elegans"	1	L29052	7653 bp	DNA	INV	23-FEB-1994
/strain="N2"	1	L29052	7653 bp	DNA	INV	23-FEB-1994
/db_xref="taxon:6239"	1	L29052	7653 bp	DNA	INV	23-FEB-1994
/gene="ced-3"	1	L29052	7653 bp	DNA	INV	23-FEB-1994
/number=1	1	L29052	7653 bp	DNA	INV	23-FEB-1994
/join(2232..2366,2430..2575,2854..3107,4303..4634,	1	L29052	7653 bp	DNA	INV	23-FEB-1994
5547..5760,5815..5942,6298..6537,7013..7075)	1	L29052	7653 bp	DNA	INV	23-FEB-1994
/gene="ced-3"	1	L29052	7653 bp	DNA	INV	23-FEB-1994
/note="codes for a protein similar to mammalian	1	L29052	7653 bp	DNA	INV	23-FEB-1994
interleukin-1b-converting enzyme"	1	L29052	7653 bp	DNA	INV	23-FEB-1994
/codon_start=1	1	L29052	7653 bp	DNA	INV	23-FEB-1994
/product="cell death protein"	1	L29052	7653 bp	DNA	INV	23-FEB-1994

Matches 66; Conservative 0; Mismatches 16; Indels 2; Gaps 2;

Db 104 GAAATGCGTATGTCAGCATATTTGACAGCAAAATATTTGTCGCGAAACTACAGA 163
 |||||
 QY 3134 GAGATGCGCATTAACAATATTTGACGCGCAAA-TATCTCGTAGCGAAATAACAGT 3192
 |||||
 Db 164 ATCCCT-AAATGACTACTAGT 186
 |||||
 QY 3193 AACCTTTAAATGACTATAGT 3216

RESULT 15

LOCUS D75308 360 bp mRNA EST 14-DEC-1995
 DEFINITION CELK098H4F Yuji Kohara unpublished cDNA Caenorhabditis elegans CDNA
 clone YK98h4 5', mRNA sequence.

ACCESSION

D75308
 g1121092

VERSION

D75308.1 GI:1121092

KEYWORDS

EST.

SOURCE

Caenorhabditis elegans.

ORGANISM

Caenorhabditis elegans.

REFERENCE

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

AUTHORS

Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
 Tabara,H.

TITLE

Toward an expression map of the C.elegans genome

JOURNAL

Unpublished (1994)

COMMENT

Contact: Yuji Kohara

Gene Library Lab

National Institute of Genetics

Yata 111, Mishima, Shizuoka 411, Japan

Tel: 0559-75-0771

Fax: 0559-75-6240

Email: ykohara@ddbj.nig.ac.jp

Insert length: 895 Std Error: 0.00

High quality sequence stop: 363.

FEATURES

source

location/Qualifiers

1..360

/organism="Caenorhabditis elegans"

/strain="CB1489 him-8(e1489)"

/note="dev_stage=varied, sex=Hermaphrodite male,"

tissue_type=whole animal"

/db_xref="taxon:6239"

/clone="YK98h4"

/clone_lib="Yuji Kohara unpublished cDNA"

107 a 81 c 52 g 119 t 1 others

BASE COUNT
 ORIGIN

Query Match

Best Local Similarity 83.1%; Score 38; DB 32; Length 360;
 Matches 54; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

Db 82 TACAGTGCATTTAAGAAATTAAGTATTTTCTACATAGCAATATTTTGACGCTCAAC 141
 |||||
 Cp 3214 TACAATAGTCATTTAAAGGGTTACTGATTTTTCGCTACGAGATATTTGCG-CGTCAAT 3156
 |||||
 Db 142 ATGTT 146
 |||||
 Cp 3155 ATGTT 3151

Search completed: Fri Aug 6 05:15:58 1999
 Job time : 1741 secs.

Db	1	TTTCGCTAAGACTACTTTGGCGGTAAGTATGTCGCCCATACGATTCACAAATTTT	60
Cp	3184	TTTCGCTACGAGATA-TTTCGCCGTCAAAATATGTTGAGTAATGCCGATTCACGAATTTA	3126
Db	61	ATGTT	65
Cp	3125	GTGTT	3121
RESULT	12		
LOCUS	C29898	300 bp	mRNA
DEFINITION	C29898	Yuji Kohara unpublished cDNA	Caenorhabditis elegans CDNA
ACCESSION	C29898	clone yk228a6 3', mRNA sequence.	
NID	g2361694		
VERSION	C29898.1	GI:2361694	
KEYWORDS	EST.		
SOURCE	Caenorhabditis elegans.		
ORGANISM	Caenorhabditis elegans		
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;		
AUTHORS	Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.		
TITLE	1 (bases 1 to 300)		
JOURNAL	Kohara,Y., Mochizashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,		
COMMENT	Sano,M., Miyata,A. and Nishigaki,A.		
	Expression map of the C.elegans genome		
	Unpublished (1996)		
	On Sep 12, 1996 this sequence version replaced gi:1393504.		
FEATURES			
source	Contact: Yuji Kohara		
	Gene Library lab		
	National Institute of Genetics		
	Yata 1111, Mishima, Shizuoka 411, Japan		
	Tel: 0559-75-0771		
	Fax: 0559-75-6240		
	Email: ykohara@dbj.nig.ac.j.		
	location/Qualifiers		
	1.300		
	/organism="Caenorhabditis elegans"		
	/strain="CB1489 him-8(el489)"		
	/note="dev_stage=varied, sex=Hermaphrodite male,		
	tissue_type=whole animal"		
	/db_xref="taxon:6239"		
	/clone="YK228a6"		
	/clone_lib="Yuji Kohara unpublished CDNA"		
	/clone_11b="42 c		
	BASE COUNT	111 a	42 c
	ORIGIN	60 g	86 t
		1 others	
	Query Match	3.8%;	Score 38; DB 15; Length 300;
	Best Local Similarity	90.4%;	Pred.No. 4.20e-13;
	Matches	47; Conservative	0; Mismatches 4; Indels 1; Gaps 1;
Db	164	AAATGACTACTGTANTTTTCGGCTACGAGATATTTGGCGGCTCAATATGTTG	215
Cp	3200	AAAGGTTACTGTATTTTTCGCTACGAGATATTT-GCGCGTCAATATGTTG	3150
RESULT	13		
LOCUS	C32017	300 bp	mRNA
DEFINITION	C32017	Yuji Kohara unpublished CDNA	Caenorhabditis elegans CDNA
ACCESSION	C32017	clone yk314a7 3', mRNA sequence.	
NID	g2363813		
VERSION	C32017.1	GI:2363813	
SOURCE	EST.		
ORGANISM	Caenorhabditis elegans.		
	Caenorhabditis elegans		
	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;		
	Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.		
	1 (bases 1 to 300)		
	Kohara,Y., Mochizashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,		
	Sano,M., Miyata,A. and Nishigaki,A.		
	Expression map of the C.elegans genome		
TITLE			

JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1404980.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@ddbj.nig.ac.jp.

FEATURES
source
1. .300
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
/db_xref="taxon:6239"
/clone="yk314a7"
/clone_lib="Yuji Kohara unpublished cDNA"

BASE COUNT 103 a 47 c 58 g 82 t 10 others
ORIGIN

Query Match 3.8%; Score 38; DB 15; Length 300;
Best Local Similarity 88.7%; Pred. No. 4.20e-13;
Matches 47; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Db 142 TAAATGACTACTGTATTTTCGCTACNAGATATTTGGCGGTCNATATGTG 194
||||| 1 ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
CP 3201 TAAAGGTTACTGTATTTTCGCTACGAGATATTT-GCGCGTCAATATATGTG 3150

RESULT	14	D64660	337 bp	mRNA	EST	13-DEC-1995
LOCUS		CELK086DZR	Yuji Kohara unpublished	CDNA	Caenorhabditis elegans	CDNA
DEFINITION		clone yk86d12 3',	mRNA sequence.			
ACCESSION		D64660				
NID		g1116350				
VERSION		D64660.1	GI:1116350			
KEYWORDS		EST.				
SOURCE		Caenorhabditis elegans.				
ORGANISM		Caenorhabditis elegans				
REFERENCE		Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.				
AUTHORS		1 (bases 1 to 337) Kohara, Y., Mitsuki, H., Nishigaki, A., Mochashiki, T., Sugimoto, A. and Tabara, H.				
TITLE		Toward an expression map of the C.elegans genome				
JOURNAL		Unpublished (1994)				
COMMENT		On Apr 14, 1993 this sequence version replaced gi:503294.				
FEATURES		source				
source		Contact: Yuji Kohara Gene Library Lab National Institute of Genetics Yata 111, Mishima, Shizuoka 411, Japan Tel: 0559-75-0771 Fax: 0559-75-6240 Email: ykohara@ddbj.nig.ac.jp High quality sequence stop: 267.				
Location/Qualifiers		1..337				
/organism="Caenorhabditis elegans"		/strain="CB1489 him-8(e1489)"				
/note="dev_stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"		/db_xref="taxon:6239"				
/map="117"		/clone="yk86d12"				
/clone_lib="Yuji Kohara unpublished CDNA"		/clone="116 a 52 c 72 g 94 t 3 others"				
BASE COUNT		116 a	52 c	72 g	94 t	3 others
ORIGIN						
Query Match		3.8%;	Score 38;	DB 32;	Length 337;	
Best Local Similarity		78.6%;	Pred. No. 4.20e-13;			

TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL Unpublished (1998)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797455.

Contact: Eun M.Y.
Department of CytoGenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.

FEATURES

source

1. 247
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone="97SN1784"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT 7 a 16 c 21 g 34 t 169 others
ORIGIN

Query Match 4.3%; Score 43; DB 17; Length 247;
Best Local Similarity 11.7%; Pred. No. 3.81e-18;
Matches 26; Conservative 103; Mismatches 92; Indels 2; Gaps 2;

Db 22 KHTHMTBWCVRVGTITNNGKHNGRTTWNDCSDNAHCRTVBYARSKYGYTBY 81
CP 3020 TTGCTCTTGAGTAGATATCTCGATAGAAGATGATGACACTGAAGAGACGCTGTAC 2961
Db 82 YSWNVDNTGCTGVKTTVNVHSGWNNCSNVVYVWBTAYCDYBHYBDRAHVDPTRCT 141
CP 2960 GGTGAACTCGGCTCGTAGAGTCGCGCGGCTCAATGCGCGCTCCGACGATGCGC 2901
Db 142 NDRGYCNYTASDNGTSATKRVTYGDKTSDCGGCGWRKVTYSSBYBRGVNVMVRTSM 201
CP 2900 TTGCCGCGTACATTGAC-ACTCGAATTTCGACAGCATTCGATCAACACTGGAAGG 2842
Db 202 WTDKSTKM-BSDMSRSRVHYGRMMBNKKRGMRSRNTDTKTW 243
CP 2841 ATATTGACGAGATGT 2799

RESULT 10
LOCUS C52036 300 bp mRNA EST 11-SEP-1997
DEFINITION C52036 Yuji Kohara unpublished cDNA Caenorhabditis elegans CDNA
clone yk225g1 3', mRNA sequence.
ACCESSION C52036
NID 92389793
VERSION C52036.1 GI:2389793
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 300)
AUTHORS Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A.,
Sano, M., Miyata, A. and Nishigaki, A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Nov 29, 1993 this sequence version replaced gi:430548.

Contact: Yuji Kohara
Gene Library Lab

National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.
Location/Qualifiers

FEATURES

source

1. 300
/organism="Caenorhabditis elegans"
/strain="CBL489 him-8(e1489)"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
/db_xref="taxon:6239"
/map="21"
/clone="yk225g1"
/clone_lib="Yuji Kohara unpublished CDNA"
BASE COUNT 115 a 44 c 52 g 87 t 2 others
ORIGIN

Query Match 3.8%; Score 38; DB 15; Length 300;
Best Local Similarity 78.8%; Pred. No. 4.20e-13;
Matches 52; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Db 194 CCGTAATCGACACAGCTCTACAGTAATTAAGAATTACTGTTTTCCTACG 253
QY 3332 CCGTAACAGACACAGCGCTACAGTACTCTTTAAAGAGTTACAGTACTTTCGCTCA 3391
Db 254 AGATAT 259
QY 3392 AGATAT 3397

RESULT 11
LOCUS C53842 300 bp mRNA EST 11-SEP-1997
DEFINITION C53842 Yuji Kohara unpublished CDNA Caenorhabditis elegans CDNA
clone yk317f5 3', mRNA sequence.
ACCESSION C53842
NID 92391599
VERSION C53842.1 GI:2391599
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 300)
AUTHORS Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A.,
Sano, M., Miyata, A. and Nishigaki, A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Apr 14, 1993 this sequence version replaced gi:716436.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.
Location/Qualifiers

FEATURES

source

1. 300
/organism="Caenorhabditis elegans"
/strain="CBL489 him-8(e1489)"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
/db_xref="taxon:6239"
/map="871E01; 5; 5q21.3"
/clone="yk317f5"
/clone_lib="Yuji Kohara unpublished CDNA"
BASE COUNT 129 a 36 c 32 g 103 t
ORIGIN

Query Match 3.8%; Score 38; DB 15; Length 300;
Best Local Similarity 83.1%; Pred. No. 4.20e-13;
Matches 54; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

ORIGIN

Query Match 6.3%; Score 63; DB 8; Length 359;
Best Local Similarity 85.4%; Pred. No. 4.74e-40;
Matches 82; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

Db 22 TACAGTAGCATTTAAAGGATTACTGATTTCGCTATGAGATATTTGCGCGTGAAT 81
Cp 3214 TACATAGTCAATTTAAAGGTTACTGATTTCGCTAGAGATATTT-GCGCGTCAAT 3156
Db 82 ATGTGTGCAATACGTATTCCTGCAATTTTGGCTTC 117
Cp 3155 ATGTGAGTAAATGCGCATTCCTCAGAAATTTAGTGTTC 3120

RESULT 2

LOCUS C51090 377 bp mRNA EST 11-SEP-1997
DEFINITION C51090 Yuji Kohara unpublished cDNA Caenorhabditis elegans CDNA
clone yk491a5 5', mRNA sequence.
ACCESSION C51090
NID 92388343
VERSION C51090.1 GI:2388343
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A.,
Sano, M., Miyata, A. and Nishigaki, A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1405104.

FEATURES
source Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.
location/Qualifiers
1..377
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
/db_xref="taxon:6239"
/map="11q23"
/clone_lib="yk491a5"
/clone_lib="Yuji Kohara unpublished CDNA"
BASE COUNT 106 a 74 c 103 g 92 t 2 others
ORIGIN

Query Match 6.2%; Score 62; DB 15; Length 377;
Best Local Similarity 100.0%; Pred. No. 6.60e-39;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 316 TGTGACTCGAATGCTGTCGAATTCGAGTGTCCAAATGTCACCGCAAGCCATCGTCGAG 375
Q7 2854 TGTGACTCGAATGCTGTCGAATTCGAGTGTCCAAATGTCACCGCAAGCCATCGTCGAG 2913
Db 376 CC 377
QY 2914 CC 2915

RESULT 3
LOCUS AA754459 252 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII CDNA Library Oryza sativa
CDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID 92801165

VERSION AA754459.1 GI:2801165
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
1 (bases 1 to 252)
Nahm, B.H., Kim, J.K., Cheong, J.J., Kim, S.I., Hahn, T.R., Moon, E.P.,
Kim, W.T., Kim, W.Y., Yang, M.S., Park, R.D., Sohn, U.I., Kang, K.Y.,
Lee, M.C. and Eun, M.Y.
Large-scale sequencing Analysis of ESTs from Rice Immature Seed
Unpublished (1998)
On Jan 14, 1998 this sequence version replaced gi:1797457.

TITLE
JOURNAL
COMMENT

FEATURES
source Contact: Eun M.Y.
Department of Cyto genetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
location/Qualifiers
1..252
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/map="6"
/clone_lib="97SN1787"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT 5 a 21 c 12 g 35 t 179 others
ORIGIN

Query Match 5.3%; Score 53; DB 17; Length 252;
Best Local Similarity 10.9%; Pred. No. 8.12e-29;
Matches 26; Conservative 119; Mismatches 91; Indels 3; Gaps 3;

Db 15 CBAMMTSYBCHGNBWWVCVASHGNYMSVHNCITBRTGCHDCCKNVNWSMTWGTWNBVS 74
QY 3351 CTACAGTACTCTTTTAA-GAGTTACAGTAGTTCCTTCAAGATATTTGAAAAAGAA 3409
Db 75 GDMHWBYBNTKVDVGNHTRCSRMBVTRMAHYHDYTNBRYNNNDYHMMHBBMYBTGC 134
QY 3410 TTTAAACATTTTGAAAAAATCATCTACATGTGCCAAACGCTTTTCAAGTTTCGC 3469
Db 135 MTCTMWCWB-HYNTKCTASGWHSTINYDVKSSTNTWGTBSYDKSMHGYWCSBBVKYHTX 193
QY 3470 AGATTTTGTGATTTTTCATTCAGATATGCTTATTACACACATATTA-TCAATTAAT 3528
Db 194 VSTTRATRSYTCVRKCYVMATTKVKKYHVBGCHBTDSKCKTMWMTNKHVMTSTFD 252
QY 3529 GTGAATTTCTGTAGAAATTTTGGGCTTTCGTTCTAGATGCTCTACTTTTGAATTCG 3587

RESULT 4
LOCUS AA754459 252 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII CDNA Library Oryza sativa
CDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID 92801165
VERSION AA754459.1 GI:2801165
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa

 M E R E F
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Fri Aug 6 04:46:57 1999; Maspar time 1724.51 Seconds
 1358.744 Million cell updates/sec

Tabular output not generated.

Title: >US-08-287-669-18
 Description: (2701-3700) from US08287669.seq (4 of 10)
 Perfect Score: 1000
 N.A. Sequence: 2701 TTGAGTGAATATATTTA.....GTAAAAAGCGCATGCTTT 3700
 Comp: AACTTCACCTTATATAAAT.....CATTTTTCGCGTACGTAAA

Scoring table: TABLE default
 Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: emb1-est58
 1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
 6:em_est9 7:em_gss1

Database: genbank-est111
 8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
 13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
 17:gb_est18 18:gb_est19 19:gb_est20 20:gb_est21
 21:gb_est22 22:gb_est23 23:gb_est24 24:gb_est25
 25:gb_est26 26:gb_est27 27:gb_est28 28:gb_est29
 29:gb_est30 30:gb_est31 31:gb_est32 32:gb_est33 33:gb_est34
 34:gb_est35 35:gb_est36 36:gb_est37 37:gb_est38 38:gb_est39
 39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 11.614; Variance 3.619; scale 3.209

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
c 1	63	6.3	359	8	T01235	WEST01956 Early embryo	4.74e-40
c 2	62	6.2	377	15	C51090	C51090 Yuji Kohara unp	6.60e-39
c 3	53	5.3	252	17	AA754459	97SN1787 Rice Immature	8.12e-29
c 4	53	5.3	252	17	AA754459	97SN1787 Rice Immature	8.12e-29
c 5	45	4.5	357	15	C41501	C41501 Yuji Kohara unp	3.19e-20
c 6	45	4.5	359	8	T01235	WEST01956 Early embryo	3.19e-20
c 7	44	4.4	357	15	C41501	C41501 Yuji Kohara unp	3.52e-19
c 8	43	4.3	247	17	AA754458	97SN1784 Rice Immature	3.81e-18
c 9	43	4.3	247	17	AA754458	97SN1784 Rice Immature	3.81e-18
c 10	38	3.8	300	15	C52036	C52036 Yuji Kohara unp	4.20e-13

c 11	38	3.8	300	15	C53842	C53842 Yuji Kohara unp	4.20e-13
c 12	38	3.8	300	15	C29898	C29898 Yuji Kohara unp	4.20e-13
c 13	38	3.8	300	15	C32017	C32017 Yuji Kohara unp	4.20e-13
c 14	38	3.8	337	32	D64660	CELK086DZR Yuji Kohara	4.20e-13
c 15	38	3.8	360	32	D75308	CELK098H4F Yuji Kohara	4.20e-13
c 16	38	3.8	2275	20	AF034173	AF034173 Human mRNA (T	4.20e-13
c 17	38	3.8	2275	20	AF034173	AF034173 Human mRNA (T	4.20e-13
c 18	38	3.7	300	15	C29898	C29898 Yuji Kohara unp	4.00e-12
c 19	37	3.7	337	32	D64660	CELK086DZR Yuji Kohara	4.00e-12
c 20	37	3.7	360	15	C39290	CELK098H4F Yuji Kohara	2.06e-07
c 21	32	3.2	360	32	D75308	CELK098H4F Yuji Kohara	1.64e-06
c 22	31	3.1	807	21	C29734	C29734 Yuji Kohara unp	1.26e-05
c 23	31	3.0	300	15	C29734	C29734 Yuji Kohara unp	1.26e-05
c 24	30	3.0	408	22	A1070726	UI-R-C2-no-a-01-0-UI.s	1.26e-05
c 25	30	3.0	448	37	AQ003622	CPG0285A CpiowagDNA1 C	9.30e-05
c 26	29	2.9	255	22	A1059267	UI-R-C1-1b-a-05-0-UI.s	9.30e-05
c 27	29	2.9	300	15	C35709	C35709 Yuji Kohara unp	9.30e-05
c 28	29	2.9	300	15	C52036	C52036 Yuji Kohara unp	9.30e-05
c 29	29	2.9	401	42	AQ445355	GSSTC01014 Trypanosoma	9.30e-05
c 30	29	2.9	429	17	AA740614	ny97f09.s1 NCI_CGAP_GC	9.30e-05
c 31	29	2.9	429	17	AA740614	ny97f09.s1 NCI_CGAP_GC	9.30e-05
c 32	29	2.9	559	26	AU005020	AU005020 Bombyx mori p	6.59e-04
c 33	29	2.9	163	30	R45014	yg36e10.s1 Soares infra	6.59e-04
c 34	28	2.8	252	21	C94106	C94106 Dictyostelium d	6.59e-04
c 35	28	2.8	430	21	AA971574	OP84C04.s1 Soares_NFL	6.59e-04
c 36	28	2.8	562	26	AU001283	AU001283 Bombyx mori p	6.59e-04
c 37	28	2.8	620	25	AU039156	AU039156 Dictyostelium	6.59e-04
c 38	28	2.8	894	10	C22780	C22780 Dictyostelium d	4.47e-03
c 39	28	2.8	164	20	C89844	C89844 Dictyostelium d	4.47e-03
c 40	27	2.7	300	15	C35137	C35137 Yuji Kohara unp	4.47e-03
c 41	27	2.7	300	9	AA216925	mt71b12.rl Soares mous	4.47e-03
c 42	27	2.7	371	12	AA390241	CPEST.231 PSKImunusCP	4.47e-03
c 43	27	2.7	410	26	AI394542	tf77d02.x1 NCI_CGAP.Br	4.47e-03
c 44	27	2.7	574	41	AQ347729	RPC111-13717.TV RPC111	4.47e-03
c 45	27	2.7					

ALIGNMENTS

RESULT 1
 LOCUS T01235 359 bp mRNA EST 10-NOV-1992
 DEFINITION WEST01956 Early embryo, Stratagene (cat. #937007) Caenorhabditis
 elegans CDNA clone CEES041, mRNA sequence.

ACCESSION T01235
 NID 9277716
 VERSION T01235.1 GI:277716
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 359)
 McCombie,W.R., Kelley,J.M., Aubin,L., Goscoechea,M.,
 Fitzgerald,M.G., Wu,A., Adams,M.D., Dubnick,M., Kerlavage,A.R.,
 Venter,J.C. and Fields,C.A.
 Caenorhabditis elegans cDNAs
 Unpublished (1993)
 Other ESTs: WEST01957
 COMMENT Contact: Kerlavage, AR

Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 Seq primer: M13 Forward.
 Location/Qualifiers

FEATURES
 source
 1. .359
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /clone="CEES041"
 /clone_lib="Early embryo, Stratagene (cat. #937007)"
 BASE COUNT 115 a 61 c 60 g 122 t 1 others

FT	/tag= p
FT	/rpt_type= INVERTED
FT	/note= "Repeat 3"
FT	5547..5760
FT	/tag= q
FT	/number= Exon_5
FT	5757
FT	/tag= i
FT	/note= "G>A, from allele n2433"
FT	5761..5814
FT	/tag= s
FT	/number= Intron_5
FT	5815..5942
FT	/tag= t
FT	/number= Exon_6
FT	5943..6297
FT	/tag= u
FT	/number= Intron_6
FT	6062..6138
FT	/tag= v
FT	/rpt_type= INVERTED
FT	6298..6537
FT	/tag= w
FT	/number= Exon_7
FT	6538..7012
FT	/tag= x
FT	/number= Intron_7
FT	6567..6625
FT	/tag= y
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 6905-6965"
FT	6905..6965
FT	/tag= z
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat 6567-6625"
FT	7012..7075
FT	/tag= aa
FT	/number= Exon_8
PN	WO9325685-A.
PD	23-DEC-1993.
PE	14-JUN-1993; U05701.
PR	12-JUN-1992; US-897788.
PR	20-NOV-1992; US-979638.
PA	(MASI) MASSACHUSETTS INST TECHNOLOGY.
PI	Horvitz HR, Shaham S, Yuan J;
DR	WPI; 94-007542/01.
DR	P-PsDB; R53281.
PT	Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT	develop agents to increase or prevent cell death in organisms
PS	Claim 14; Fig 4; 127pp; English.
CC	The sequences given in Q64735-45 represent mutations of the C. elegans
CC	ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC	was most abundant in embryos, but was also detected in larvae and young
CC	adults, suggesting that ced-3 is expressed not only in cells undergoing
CC	cell death. The four largest introns as well as sequences 5' of the
CC	start codon contain repetitive elements, some of which have been
CC	characterised in non-coding regions of other C. elegans genes, such
CC	as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC	length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC	region that might be a transmembrane region. One region of Ced-3 is
CC	very rich in serine. It is thought that this region is involved in
CC	protein-protein interactions, similar to acid blobs in transcription
CC	factors. Of the mutations which occur within the ced-3 gene, eight of
CC	the mutations are missense mutations, two are nonsense mutations and
CC	two are putative splicing mutations. These mutations establish the
CC	null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC	function is not essential for viability. The ced-3 and ced-4 gene
CC	products may be used to develop agents for treating conditions
CC	characterised by cell deaths, such as myocardial infarction, stroke,
CC	degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC	infection, aging or hair loss.
SQ	Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

Query Match	99.6%;	Score 996;	DB 9;	Length 7653;
Best Local Similarity	99.8%;	Pred. No. 0.00e+00;		
Matches 998;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
Db 2701	ttgaagtgaatatatttatttactgaaagctcgagtattattttttaacacta	2760		
QY 2701	TTGAAGTGAATAATATTTTATTACTGAAAGCTCGAGTGATTAATTTTAAACACTA	2760		
Db 2761	atttcgtgcgcaaaagcgccatttgragatttgcgaaataacttgtcacacacac	2820		
QY 2761	ATTTTCGTGGCGCAAAAGGCCATTTTGTAGATTGGCCGAAATACTTGTCCACACACAC	2820		
Db 2821	acacacatctccttcaatatatcccttttccagtgltgactcgaaatgtctcgaaatcga	2880		
QY 2821	ACACACATCTCCTTCAATATATCCCTTTTCCAGTGTGACTCGAATGCTGTGGAATTGCA	2880		
Db 2881	gtgtccaatgtcacccgcaagccatcgttcggaagccgcgcatltyagcccccggtacac	2940		
QY 2881	GTGTCCAATGTCAACCGGCAAGCCATCTCGGAGCCGCCGATTGAGCCCCCGGCTACAC	2940		
Db 2941	ttcacccgacccgagttcacccgtgacagcgctccttcagtgtcaatcattcacttcatca	3000		
QY 2941	TTCACCGACCCGAGTTCAACCGTGACAGCGTCTCTTCAGTGTCACTTCACTTCTTATCA	3000		
Db 3001	ggatatctactcaagaagcaagatcgtlctcgatcgcgtgcactlcatcatcgatcg	3060		
QY 3001	GGATATCTACTCAAGAGCAAGATCTGTTCTCGATCGCGTGCACCTTCATTCATCGGATCG	3060		
Db 3061	acacaattattcatctcctccagtcacagcatlccccagccaacctgtatgttgaatcg	3120		
QY 3061	ACACAATTATTCACTCTCCCTCCAGTCAACGCATTTCGCCAACCTTGTATGTTGATGCG	3120		
Db 3121	aacactaaatlctgagaatgcgcatlactcaacatatattgacgcgcaaatatctcgtagc	3180		
QY 3121	AACACTAAATTCTGAGAATGGCATTACTCAACATATTTGACGGCGCAATATCTCGTAGC	3180		
Db 3181	gaaaaatacaglaacccttaaatgtactatgttagtgcattlaagcgctcgatttgcg	3240		
QY 3181	GAAAAATACAGTAACCCCTTAATGACTATTGTAGTGTGATTTACGGGCTCGAATTTTCG	3240		
Db 3241	aaacgaatatatgctcgaatttgcacaacgaattttaaatttgcatttltgtgtttcct	3300		
QY 3241	AAACGAATATATGCTCGAATTGTGACACAGCAATTTTAATTTGTCAATTTTGTGTTTCTT	3300		
Db 3301	ttgatatttlttgatccaatataaatlatlccgtaaaacagacaccagcggtacagtlact	3360		
QY 3301	TTGATATTTTGTATCAATTAATAATTAATTTCCGTAAACAGACACACAGCGCTACAGTACT	3360		
Db 3361	cttttaagaagttacagtagtlltgcgtlccaagatatllgaaaagaattttaaacattt	3420		
QY 3361	CTTTTAAAGAGTTACAGTAGTTTTTCGCTTCAAGATATTTTGAAGAAATTTTAAACATTT	3420		
Db 3421	tgaaaaaaaaatcatctaacatgtgccaaaacgctlltccaaglttcgcagatttttga	3480		
QY 3421	TGAAAAAAAATCATCTAACATGTGCCAAAACGCTTTTTCAGAGTTTCGAGATTTTGTGA	3480		
Db 3481	tttttccaatccaagatatgcttattaacacatatataatcatctaagtgaatttcttg	3540		
QY 3481	TTTTTTCAATCAAGATATGCTTATTAAACACATATAATATCATTAATGTGAATTTCTTG	3540		
Db 3541	tagaaaatttggcgcttctcgtlctagtagtcttacttltgaatitgtccaacgaaaaa	3600		
QY 3541	TAGAAATTTTGGCGTTTTCGTTCTAGTAGTCTCTACTTTTGAATTTGCTCAACGAAAAA	3600		
Db 3601	tcatgtgttggttcataatgaatgacaaaaaataagcaatttttataatatlttccctat	3660		
QY 3601	TCATGTGTTGTTCATATGAATGACGAAAAATAGCAATTTTATATATTTTCCCTAT	3660		
Db 3661	tcatgttggtgcagaaaaatagtaaaaaagcgcatgcatl 3700			
QY 3661	TCATGTGTGTCAGAAAAATAGTAAAAAAGCGCATGCAATTT 3700			

CC splicing mutations (see also Q64735-45). These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2429 A; 1453 C; 1272 G; 2499 T;

Query Match 99.6%; Score 996; DB 9; Length 7653;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 2701 ttgaagtgaatatattttatttacttgaagctcgagtgattattattttaacacta 2760
|||
Qy 2701 ttgaagtgaatatattttatttacttgaagctcgagtgattattattttaacacta 2760
Db 2761 atttcgtgagcgaagccattttgtagatttcgcgaaataacttgcacacacacac 2820
|||
Qy 2761 atttcgtgagcgaagccattttgtagatttcgcgaaataacttgcacacacacac 2820
|||
Db 2821 acacacatccttcaaatatcccttttccagtggtgactcgaatgctgcgaattcga 2880
|||
Qy 2821 acacacatccttcaaatatcccttttccagtggtgactcgaatgctgcgaattcga 2880
|||
Db 2881 gtgtccaatgtcacccgcaagccatcgtcgagccgcgcatltagcccccgcgcgtacac 2940
|||
Qy 2881 gtgtccaatgtcacccgcaagccatcgtcgagccgcgcatltagcccccgcgcgtacac 2940
|||
Db 2941 ttcacgcgacccgagttcacccgtgacagcgtctcttcagtgatcatcattcattatca 3000
|||
Qy 2941 ttcacgcgacccgagttcacccgtgacagcgtctcttcagtgatcatcattcattatca 3000
|||
Db 3001 ggatatctactcaagagcaagatctcgttctcgtatcgcgctcattcatcgcgacgcg 3060
|||
Qy 3001 ggatatctactcaagagcaagatctcgttctcgtatcgcgctcattcatcgcgacgcg 3060
|||
Db 3061 acacaattatcatctcctccagtcacagcatttccacgcaacctgtatgttagtcg 3120
|||
Qy 3061 acacaattatcatctcctccagtcacagcatttccacgcaacctgtatgttagtcg 3120
|||
Db 3121 aacactaaatctgagaaatgctgcgcatlactcaacatatatttgacgcgcaaatatctcgtagc 3180
|||
Qy 3121 aacactaaatctgagaaatgctgcgcatlactcaacatatatttgacgcgcaaatatctcgtagc 3180
|||
Db 3181 gaaaaatacagtaaccctttaaagtactatgtgtagtgcgatttacggtcgcgatttcg 3240
|||
Qy 3181 gaaaaatacagtaaccctttaaagtactatgtgtagtgcgatttacggtcgcgatttcg 3240
|||
Db 3241 aaacgaatatatgctcgaatgtgacaacgaattttaaattgtcatttttggtttctt 3300
|||
Qy 3241 aaacgaatatatgctcgaatgtgacaacgaattttaaattgtcatttttggtttctt 3300
|||
Db 3301 ttgatatcttgatcaatataataattatcttcglaaacagacacagcggtacagttact 3360
|||
Qy 3301 ttgatatcttgatcaatataataattatcttcglaaacagacacagcggtacagttact 3360
|||
Db 3361 ctttaagaggtacagtagtttcgcctcaagatatattgaaagaattttaaacattt 3420
|||
Qy 3361 ctttaagaggtacagtagtttcgcctcaagatatattgaaagaattttaaacattt 3420
|||
Db 3421 tgaataaaaaaatcatcctaactgtgccaaaacgctttttcaagttcgcgagatttttga 3480
|||
Qy 3421 tgaataaaaaaatcatcctaactgtgccaaaacgctttttcaagttcgcgagatttttga 3480
|||
Db 3481 ttttttcatcgaagatgcttatttaacacataataattatcatatgtaatttcttg 3540
|||
Qy 3481 ttttttcatcgaagatgcttatttaacacataataattatcatatgtaatttcttg 3540
|||
Db 3541 tagaaatttggtgcttctctagtagtcttactcttggaaatgctcacaacgaaaaa 3600
|||
Qy 3541 tagaaatttggtgcttctctagtagtcttactcttggaaatgctcacaacgaaaaa 3600
|||

Db 3601 tcatgtgtttgttcatatgaatgaccaaataagcaattttatatatatccctat 3660
|||
Qy 3601 tcatgtgtttgttcatatgaatgaccaaataagcaattttatatatatccctat 3660
|||
Db 3661 tcatgtgtgcaaaaaaatagtaaaaaaagcgcatgattt 3700
|||
Qy 3661 tcatgtgtgcaaaaaaatagtaaaaaaagcgcatgattt 3700
|||
RESULT 13
ID 064737 standard; DNA; 7653 BP.
AC 064737;
DE 23-JUN-1994 (first entry)
KW ced-3 (G5757A) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key location/Qualifiers
FT repeat_unit 1356..1472
FT FT /tag= a
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT FT 1490..1614
FT FT /tag= b
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT FT 2232..2366
FT FT /tag= c
FT FT /number= Exon_1
FT FT 2367..2429
FT FT /tag= d
FT FT /number= Intron_1
FT FT 2450..2575
FT FT /tag= e
FT FT /number= Exon_2
FT FT 2576..2853
FT FT /tag= f
FT FT /number= Intron_2
FT FT 2854..3107
FT FT /tag= g
FT FT /number= Exon_3
FT FT 3108..4302
FT FT /tag= h
FT FT /number= Intron_3
FT FT 3126..3243
FT FT /tag= i
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT FT 3329..3396
FT FT /tag= j
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT FT 3487..3759
FT FT /tag= k
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT FT 3782..4070
FT FT /tag= l
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 3487-3759"
FT FT 4303..4634
FT FT /tag= m
FT FT /number= Exon_4
FT FT 4635..5546
FT FT /tag= n
FT FT /number= Intron_4
FT FT 4688..4719
FT FT /tag= o
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5221..5330
FT FT repeat_unit


```
FT FT      /*tag= a
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat at 1490-1614"
FT FT      /*tag= b
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat at 1356-1472"
FT FT      /*tag= c
FT FT      /number= Exon_1
FT FT      2310
FT FT      /*tag= d
FT FT      /note= "C>T, from allele n1040"
FT FT      2367..2429
FT FT      /*tag= e
FT FT      /number= Intron_1
FT FT      2430..2575
FT FT      /*tag= f
FT FT      /number= Exon_2
FT FT      2487
FT FT      /*tag= g
FT FT      /note= "G>A, from allele n718"
FT FT      2576..2853
FT FT      /*tag= h
FT FT      /number= Intron_2
FT FT      2854..3107
FT FT      /*tag= i
FT FT      /number= Exon_3
FT FT      3108..4302
FT FT      /*tag= j
FT FT      /number= Intron_3
FT FT      3126..3243
FT FT      /*tag= k
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat at 3329-3396"
FT FT      3329..3396
FT FT      /*tag= l
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat at 3126-3243"
FT FT      3487..3759
FT FT      /*tag= m
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat at 3782-4070"
FT FT      3782..4070
FT FT      /*tag= n
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat 3487-3759"
FT FT      4303..4634
FT FT      /*tag= o
FT FT      /number= Exon_4
FT FT      4635..5546
FT FT      /*tag= p
FT FT      /number= Intron_4
FT FT      4688..4719
FT FT      /*tag= q
FT FT      /rpt_type= INVERTED
FT FT      /note= "Repeat 3"
FT FT      5221..5330
FT FT      /*tag= r
FT FT      /rpt_type= INVERTED
FT FT      /note= "Repeat 3"
FT FT      5547..5760
FT FT      /*tag= s
FT FT      /number= Exon_5
FT FT      5757
FT FT      /*tag= t
FT FT      /note= "G>A, from allele n2433"
FT FT      5761..5814
FT FT      /*tag= u
FT FT      /number= Intron_5
FT FT      5815..5942
FT FT      /*tag= v
FT FT      /number= Exon_6
```

```
FT FT      mutation
FT FT      5940
FT FT      /*tag= w
FT FT      /note= "C>T, from allele n1165"
FT FT      5943..6297
FT FT      /*tag= x
FT FT      /number= Intron_6
FT FT      6062..6138
FT FT      /*tag= y
FT FT      /rpt_type= INVERTED
FT FT      6297
FT FT      /*tag= z
FT FT      /note= "G>A, from allele n717"
FT FT      6298..6537
FT FT      /*tag= aa
FT FT      /number= Exon_7
FT FT      6322
FT FT      /*tag= ab
FT FT      /note= "C>T, from allele n1949"
FT FT      6372
FT FT      /*tag= ac
FT FT      /note= "G>A, from allele n1286"
FT FT      6434
FT FT      /*tag= ad
FT FT      /note= "C>T, from alleles n1129 and n1164"
FT FT      6485
FT FT      /*tag= ae
FT FT      /note= "C>T, from allele n2430"
FT FT      6535
FT FT      /*tag= af
FT FT      /note= "G>A, from allele n2426"
FT FT      6538..7012
FT FT      /*tag= ag
FT FT      /number= Intron_7
FT FT      6567..6625
FT FT      /*tag= ah
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT FT      6905..6965
FT FT      /*tag= ai
FT FT      /rpt_type= INVERTED
FT FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT FT      7013..7075
FT FT      /*tag= aj
FT FT      /number= Exon_8
FT FT      7020
FT FT      /*tag= ak
FT FT      /note= "C>T, from allele n1163"
FT FT      mutation
FT FT      WO9325685-A.
FT FT      23-DEC-1993.
FT FT      14-JUN-1993; U05701.
FT FT      12-JUN-1992; US-8977788.
FT FT      20-NOV-1992; US-979638.
FT FT      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT FT      Horvitz HR, Shaham S, Yuan J;
FT FT      WPI; 94-007542/01.
FT FT      P-PSDB; R47466.
FT FT      Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT FT      develop agents to increase or prevent cell death in organisms
FT FT      Claim 2; Fig 4; 127pp; English.
FT FT      This sequence represents the C. elegans ced-3 gene. A 2.8 kb mRNA
FT FT      was identified as the ced-3 transcript and was most abundant in
FT FT      embryos, but was also detected in larvae and young adults, suggesting
FT FT      that ced-3 is expressed not only in cells undergoing cell death. The
FT FT      four largest introns as well as sequences 5' of the start codon
FT FT      contain repetitive elements, some of which have been characterised
FT FT      in non-coding regions of other C. elegans genes, such as fem-1, lin-12
FT FT      and myoD. The Ced-3 protein is 503 amino acids in length. Ced-3 is
FT FT      highly hydrophilic with no significant hydrophobic region that might
FT FT      be a transmembrane region. One region of Ced-3 is very rich in serine.
FT FT      It is thought that this region is involved in protein-protein inter-
FT FT      actions, similar to acid blobs in transcription factors. Of the
FT FT      mutations which occur within the ced-3 gene, eight of the mutations
FT FT      are missense mutations, two are nonsense mutations and two are putative
```



```
FT      intron      5943..6297
FT      /tag= t
FT      /number= Intron_6
FT      repeat_region 6062..6138
FT      /tag= u
FT      /rpt_type= INVERTED
FT      exon        6298..6537
FT      /tag= v
FT      /number= Exon_7
FT      mutation    6322
FT      /tag= w
FT      /note= "C>T, from allele n1949"
FT      intron      6538..7012
FT      /tag= x
FT      /number= Intron_7
FT      repeat_unit 6567..6625
FT      /tag= y
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT      repeat_unit 6905..6965
FT      /tag= z
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      exon        7012..7075
FT      /tag= aa
FT      /number= Exon_8
FT      WO9325685-A.
FT      23-DEC-1993.
FT      14-JUN-1993; U05701.
FT      12-JUN-1992; US-897788.
FT      20-NOV-1992; US-979638.
FT      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT      Horvitz HR, Shiham S, Yuan J;
FT      WPI; 94-007542/01.
FT      P-PSDB; R53283.
FT      Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT      develop agents to increase or prevent cell death in organisms
FT      Claim 14; Fig 4; 127pp; English.
FT      The sequences given in Q64735-45 represent mutations of the C. elegans
FT      ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT      was most abundant in embryos, but was also detected in larvae and young
FT      adults, suggesting that ced-3 is expressed not only in cells undergoing
FT      cell death. The four largest introns as well as sequences 5' of the
FT      start codon contain repetitive elements, some of which have been
FT      characterised in non-coding regions of other C. elegans genes, such
FT      as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
FT      length. Ced-3 is highly hydrophilic with no significant hydrophobic
FT      region that might be a transmembrane region. One region of Ced-3 is
FT      very rich in serine. It is thought that this region is involved in
FT      protein-protein interactions, similar to acid blobs in transcription
FT      factors. Of the mutations which occur within the ced-3 gene, eight of
FT      the mutations are missense mutations, two are nonsense mutations and
FT      two are putative splicing mutations. These mutations establish the
FT      null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
FT      function is not essential for viability. The ced-3 and ced-4 gene
FT      products may be used to develop agents for treating conditions
FT      characterised by cell deaths, such as myocardial infarction, stroke,
FT      degenerative disease, traumatic brain injury, hypoxia, pathogenic
FT      infection, aging or hair loss.
FT      Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;
SQ
Query Match 99.6%; Score 996; DB 9; Length 7653;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Db      2821 acacacatctcctcaaatatcccttttccagtggtgactcgaatgctgcaattcga 2880
QY      2821 ACACACATCTCCTCAAAATATCCCTTTTCCAGTGTGACTCGAATGCTGCAATTCGA 2880
Db      2881 gtgtccaatgtcaaccgccaagccatcgtagcgccgcatgagcccccgccgtacac 2940
QY      2881 GTGTCCAATGTCAACCGGCAAGCCATCGTAGAGCCGCGCATGAGCCCCCGGCTACAC 2940
Db      2941 ttcaaccgaccgaggttcacccgtgacagcgctcttcagtgatcatcatcacttctatca 3000
QY      2941 TTCACCGACCCGAGTTCACCCGTGACAGCGCTCTTCAAGTGCATTCATCTTATCA 3000
Db      3001 ggatactactcaagaagcaagatcgtttctcgatcgcggtgcaattcatcagatcg 3060
QY      3001 GGATATCTACTCAAGACCAAGATCTCGTTCGATCGCGGTGCATTCATTCGATCG 3060
Db      3061 acacaattatcatctcctccagtcacgcatlcccaagccaacctgtatgtatgacg 3120
QY      3061 ACACAATTATTCATCTCCTCCAGTCAACGCAATTTCCAGCCAACTTGATGTGATCG 3120
Db      3121 aacactaaattctgagaatgcgcatctactcaacatatttgacgcgcaaatatctcgtacg 3180
QY      3121 AACACTAAATTCGAGAAATGCGCATTAACAATATTGACGCGCAATATCTCGTAGC 3180
Db      3181 gaaaaatacagtaacccttaaatgactatgtgtagtgcgatttaagggctcgatttcg 3240
QY      3181 GAAAAATACAGTAACCCCTTAAATGACTAATGTAGTGCATTTACGGCTCGATTTTCG 3240
Db      3241 aaacgaatatatgctcgaattgtgacaacgaatttaattgtcatatttggtttct 3300
QY      3241 AAACGAATATATGCTCGAATGTGACACGAATTTAATTTGTGATTTGTGTTCTT 3300
Db      3301 ttgatatcttgatcaatataaatatttccgtaaacagacaccgaggtacagttact 3360
QY      3301 TTGATATTTTGTATCATTAATAATTAATTTCCGTAACAGACACACGCGCTACAGTACT 3360
Db      3361 ctttaagaggtacagtagtttgcgtcgaagatatlttgaaaagaattttaacatt 3420
QY      3361 CTTTAAAGAGTTACAGTAGTTTTCGCTTCAAGATATTTTGAAAAGAAATTTTAAACATTT 3420
Db      3421 tgaaaaaaatcatcataatgctgcaaaacgcttttcaagttcgcaagatttttga 3480
QY      3421 TGAAAAAAATCATCTACATGTGCCAAAAACGCTTTTCAAGTTTCGAGATTTTGTGA 3480
Db      3481 ttttttcattcaagatgcttatataacacataattatcatatgtaattcttg 3540
QY      3481 TTTTTCATTCACAGATATGCTTATTAACACATATAATTATCATTAATGTAATTTCTTG 3540
Db      3541 tagaaatttgggcttttcgttctagtatgctctacttttgaattgtcacaagaaaaa 3600
QY      3541 TAGAAATTTTGGGCTTTTCGTTCTAGTATGCTCTACTTTTGAATTTGCTACAGAAAAA 3600
Db      3601 tcatgtggttcttcatatgaaatgaccaaataatagcaatttttatatatlttccctat 3660
QY      3601 TCATGTGTTTGTTCATATGCAATGACGAAAAATAGCAATTTTATATATTTTCCCTAT 3660
Db      3661 tcatgtgtgcagaaaaaatagtaaaaaagcgcatgattt 3700
QY      3661 TCATGTGTGCAGAAAAAATAGTAAAAAAGCGCATGCAATT 3700
```

```
RESULT 12
ID      Q54666 standard; DNA; 7653 BP.
AC      Q54666;
DT      23-JUN-1994 (first entry)
DE      ced-3 gene.
KW      C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW      embryogenesis; cell death; hydrophilic; transmembrane; region;
KW      hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW      protein synthesis; null phenotype; calcium-binding domain; ss.
OS      Caenorhabditis elegans.
FH      Key location/Qualifiers
FT      repeat_unit 1356..1472
```


Db 2701 ttgaagtgaataataatttattactgaaagctcgagtgattattttttaacacta 2760
|||||
QY 2701 TTGAAGTGAATATATTTTATTACTGAAAGCTCGAGTGATTTATTTTAAACACTA 2760
Db 2761 atttcgtgcgcaaaaggccatttgtagatttgccgaaaaataacttgcacacacacac 2820
|||||
QY 2761 ATTTTGTGGCGCAAAAGGCCATTGTGTAGATTGGCGAAATACTTGTGCACACACACAC 2820
Db 2821 acacacatctccttcaaatatccccctttccagtggtgactcgaatgctgtcgaatcga 2880
|||||
QY 2821 ACACACATCTCCTTCAATATATCCCTTTCCAGTGTGACTCGAATGCTGTGGAATTGCA 2880
Db 2881 gtgtccaatgtcacccgcaagccatcgtcgagccgcgcatgagcccccgcgctacac 2940
|||||
QY 2881 GTGTCCAATGTCAACCGCAAGCCATCGTCGAGCCGCGCATGAGCCCCGCGGCTACAC 2940
Db 2941 ttcacgaccccgagttcacccgtgacagcgctctcttcagtgatcatcattcattctatca 3000
|||||
QY 2941 TTCACCGACCCGAGTTCACCGTGACAGCGTCTCTCAGTGTCATTCATCTCTTATCA 3000
Db 3001 ggatatctactcaagaagcaagatctcggttctcgatcgcggtgcaattcattcatcgagtcg 3060
|||||
QY 3001 GGATATCTACTCAAGAGCAAGATCTCGTCTCGATCGCGTGCACTTCATTCATCGGATCG 3060
Db 3061 acacaattatcatctcctccagtcacagcattccccagccaacctgtatgttatgctg 3120
|||||
QY 3061 ACACAATTATTCATCTCCTCCAGTCAACGCAATTTCCAGCCCAACCTTGATGTGATGCG 3120
Db 3121 aacactaaatctcgaaatgcgcatctacataatltgacgcgcaaatatctcgtagc 3180
|||||
QY 3121 AACACTAAATCTGAGAAATGCGCATTTACTCAACATATTTGACGCGCAATATCTCGTAGC 3180
Db 3181 gaaaaatacagtaaccctttaaatgactatgttagtgcattacgggctcgatttcg 3240
|||||
QY 3181 GAAAAATACAGTAACCCCTTAAATGACATTTGTAGTGTGATTTACGGGCTCGATTTTCG 3240
Db 3241 aaacgaatatatgctcgaaatgtgacacgaattttaattgtcattttgtgttctt 3300
|||||
QY 3241 AAACGAATATATGCTCGAATTTGTGACACGCAATTTTAATTTGTCAATTTTGTGTTTCT 3300
Db 3301 ttgatatttttgatcaatataatattttccgtaaacagacacccagcggtacagtact 3360
|||||
QY 3301 TTGATATTTTGTGATCAATTAATAATTTATTCGTAACAGACACACCGCTACAGTACT 3360
Db 3361 cttttaaagaagttacagtagtttctcgcttcaagatatatttgaaaaagaattttaacattt 3420
|||||
QY 3361 CTTTAAAGAGTTACAGTAGTTTTCGCTTCAAGATATTTTGAAGAAATTTTAAACATTT 3420
Db 3421 tgaaaaaaaatcatctaacaatgtgccaacacgcttttttccaagtttcgcagatttttga 3480
|||||
QY 3421 TGAATAAAATCATCTACATGTGCCAAAAACGCTTTTTCAGATTTCGAGATTTTGA 3480
Db 3481 ttttttcatccaagatatgcttataaacacataaattatcatlaatgtgaatttcttg 3540
|||||
QY 3481 TTTTTCATTCAGATATAGCTTATTAAACACATATTAATCATTAATGTGAATTTCTTG 3540
Db 3541 tagaaatttgggcttccgttctagatagctctacttttgaattgctcaacgaaaaaa 3600
|||||
QY 3541 TAGAAATTTTGGGCTTTCGTTCTAGTATGCTCTACTTTTGAATTTGCTCAACGAAAAAA 3600
Db 3601 tcatgtggttgcataatgaatgcaacaaatagcaatttttatatatatttccctat 3660
|||||
QY 3601 TCATGTGTTTGTTCATATGAAATGACGAAAAATAGCAATTTTATATATATTTCCCTAT 3660
Db 3661 tcatgtgtgacgaaaaatagtaaaaaagcgcatgcat 3700
|||||
QY 3661 TCATGTTGTGCAGAAAAATAGTAAAAAAGCGCATGCAATT 3700

RESULT 11
ID Q64740 standard; DNA; 7653 BP.
AC Q64740;

DE 23-JUN-1994 (first entry)
KW ced-3 (C6322P) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key location/Qualifiers
FT repeat_unit 1356..1472
FT /tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT 1490..1614
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT 2232..2366
FT /tag= c
FT /number= Exon_1
FT 2367..2429
FT /tag= d
FT /number= Intron_1
FT 2450..2575
FT /tag= e
FT /number= Exon_2
FT 2576..2853
FT /tag= f
FT /number= Intron_2
FT 2854..3107
FT /tag= g
FT /number= Exon_3
FT 3108..4302
FT /tag= h
FT /number= Intron_3
FT 3126..3243
FT /tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT 3329..3396
FT /tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT 3487..3759
FT /tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT 3782..4070
FT /tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT 4303..4634
FT /tag= m
FT /number= Exon_4
FT 4635..5546
FT /tag= n
FT /number= Intron_4
FT 4688..4719
FT /tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5221..5330
FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5547..5760
FT /tag= q
FT /number= Exon_5
FT 5761..5814
FT /tag= r
FT /number= Intron_5
FT 5815..5942
FT /tag= s
FT /number= Exon_6

Db 3661 tcatgttgcagaaaaatagtaaaaaagcgcattc 3700
 QY 3661 TCATGTTGTCAGAAAAATAGTAAAAAGCGCATGCAATT 3700

```

RESULT 10
ID 064743 standard; DNA; 7653 BP.
AC 064743;
DE 23-JUN-1994 (first entry)
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH key Location/Qualifiers
FT repeat_unit 1356..1472
FT FT /tag= a
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t repeat at 1490-1614"
FT FT 1490..1614
FT FT /tag= b
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT FT 2232..2366
FT FT /tag= c
FT FT /number= Exon_1
FT FT 2367..2429
FT FT /tag= d
FT FT /number= Intron_1
FT FT 2450..2575
FT FT /tag= e
FT FT /number= Exon_2
FT FT 2576..2853
FT FT /tag= f
FT FT /number= Intron_2
FT FT 2854..3107
FT FT /tag= g
FT FT /number= Exon_3
FT FT 3108..4302
FT FT /tag= h
FT FT /number= Intron_3
FT FT 3126..3243
FT FT /tag= i
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT FT 3329..3396
FT FT /tag= j
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT FT 3487..3759
FT FT /tag= k
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT FT 3782..4070
FT FT /tag= l
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 3487-3759"
FT FT 4303..4634
FT FT /tag= m
FT FT /number= Exon_4
FT FT 4635..5546
FT FT /tag= n
FT FT /number= Intron_4
FT FT 4688..4719
FT FT /tag= o
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5221..5330
FT FT /tag= p
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT repeat_unit
  
```

```

FT FT exon 5547..5760
FT FT /tag= q
FT FT /number= Exon_5
FT FT 5761..5814
FT FT /tag= r
FT FT /number= Intron_5
FT FT 5815..5942
FT FT /tag= s
FT FT /number= Exon_6
FT FT 5943..6297
FT FT /tag= t
FT FT /number= Intron_6
FT FT 6062..6138
FT FT /tag= u
FT FT /rpt_type= INVERTED
FT FT 6298..6537
FT FT /tag= v
FT FT /number= Exon_7
FT FT 6485
FT FT /tag= w
FT FT /note= "C>T, from allele n2430"
FT FT 6538..7012
FT FT /tag= x
FT FT /number= Intron_7
FT FT 6567..6625
FT FT /tag= y
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT FT 6905..6965
FT FT /tag= z
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 6567-6625"
FT FT 7012..7075
FT FT /tag= aa
FT FT /number= Exon_8
  
```

WO9325685-A.
 23-DEC-1993.
 PD 14-JUN-1993; U05701.
 PF 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-979638.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shahan S, Yuan J;
 DR WPI; 94-007542/01.
 DR P-PSDB; R53286.
 PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
 PT develop agents to increase or prevent cell death in organisms
 PS Claim 14; Fig 4; 127pp; English.
 CC The sequences given in Q64735-45 represent mutations of the C. elegans
 CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
 CC was most abundant in embryos, but was also detected in larvae and young
 CC adults, suggesting that ced-3 is expressed not only in cells undergoing
 CC cell death. The four largest introns as well as sequences 5' of the
 CC start codon contain repetitive elements, some of which have been
 CC characterised in non-coding regions of other C. elegans genes, such
 CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
 CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
 CC region that might be a transmembrane region. One region of Ced-3 is
 CC very rich in serine. It is thought that this region is involved in
 CC protein-protein interactions, similar to acid blobs in transcription
 CC factors. Of the mutations which occur within the ced-3 gene, eight of
 CC the mutations are missense mutations, two are nonsense mutations and
 CC two are putative splicing mutations. These mutations establish the
 CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
 CC function is not essential for viability. The ced-3 and ced-4 gene
 CC products may be used to develop agents for treating conditions
 CC characterised by cell deaths, such as myocardial infarction, stroke,
 CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
 CC infection, aging or hair loss.
 SQ Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.6%; Score 996; DB 9; Length 7653;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

FT /number= Intron_4
FT repeat_unit 4688..4719
FT /tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT repeat_unit 5221..5330
FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon 5547..5760
FT /tag= q
FT /number= Exon_5
FT intron 5761..5814
FT /tag= r
FT /number= Intron_5
FT exon 5815..5942
FT /tag= s
FT /number= Exon_6
FT intron 5943..6297
FT /tag= t
FT repeat_region 6062..6138
FT /tag= u
FT /rpt_type= INVERTED
FT exon 6298..6537
FT /tag= v
FT /number= Exon_7
FT mutation 6434
FT /tag= w
FT /note= "C>T, from allele n1129 and n1164"
FT intron 6538..7012
FT /tag= x
FT /number= Intron_7
FT repeat_unit 6567..6625
FT /tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7012..7075
FT /tag= aa
FT /number= Exon_8
PN WO9325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.
DR P-PSDB; R53285.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127pp; English.
CC The sequences given in Q64735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene

CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;
Query Match 99.6%; Score 996; DB 9; Length 7653;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 2701 ttgaagtgaatatatttattactgaagctcgagtgattatttttaacacta 2760
QY 2701 TTGAAGTGAATATATTTTATTACTGAAGCTCGAGTGATTATTATTTAACACTA 2760
Db 2761 atttcgtgqcgcaaaagccattttagatttgcggaataacttgtcacacacacac 2820
QY 2761 ATTTTCGTGGCGCAAAAGCCATTGTGTAGATTTCGCCGAAATACTTGTACACACACAC 2820
Db 2821 acacacatctccttcaatatcccttttccagtgttgactcgaatgctgctgaatcga 2880
QY 2821 ACACACATCTCCTTCAATATCCCTTTTCCAGTGTGACTCGAATGCTGTGCAATTCGA 2880
Db 2881 gtgtccaatgtcaccgcaagccatcgtcggagccgagcattgagcccgccgctacac 2940
QY 2881 GTGTCCAATGTCAACCGCAAGCCATCGTCGGAGCCGCCGATTGAGCCCGCGCTACAC 2940
Db 2941 ttccacgacccgagttcaccgctgacagcgtctcttcagtgatcatcattcattctatca 3000
QY 2941 TTCACGACCCGAGTTCACCGTGACAGCGTCTCTTCAAGTGTATCATTCACCTTCTATCA 3000
Db 3001 ggatatctactcaagcaagcaagatcgtctcgtatcgtcgatcgctgacattcatcgatcg 3060
QY 3001 GGATATCTACTCAAGACGAAGATCTCGTCTCGATCGCGTGCACTTCATTCATCGGATCG 3060
Db 3061 acacaattattcatctcctccagtcacagcatttcccgcccaacctgtatgttgatcg 3120
QY 3061 ACACAATTATTCTCTCTCCAGTCAACGCATTTCGCCGCCAACCTGTATGTGATGCG 3120
Db 3121 aacactaattctgagaatgcgcaattactcaacatatgtgaacgcaaatatctgtagc 3180
QY 3121 AACACTAAATTCTGGAATGCGCAATTACTCAACATATTGACGCGCAATATCTGTAGC 3180
Db 3181 gaaaaatacagtaaccctttaaactatgttagtgcgatttacgggctcgatttcg 3240
QY 3181 GAAAAATACGTAACCCCTTAAATGACTATTGTAGTGCATTACGGGCTCGATTTCG 3240
Db 3241 aaacgaatatatgctgaattggaacaagcaattttaaattgtcatttttgttttct 3300
QY 3241 AAACGAATATATGCTCGAATGTGACAACGAATTTTAATTGTATTTGTGTTCTT 3300
Db 3301 ttgatatttttgatcaattaataattatttccgtaaacagacaccagcggtacagtact 3360
QY 3301 TTGATATTTTGTATCAATTAATAATTATTTCGTAACAGACACACAGCGGCTACAGTACT 3360
Db 3361 ctttaagagttacagtagtttgcgtctcaagatatlttgaagaattttaaacttt 3420
QY 3361 CTTTAAAGAGTTACAGTAGTTTTCGCTTCAAGATATTTTGAAAAAAGATTTAACATTT 3420
Db 3421 tgaaaaaaatcatctaacaatgtgcaaaacgcttlttcaagtttcgcagatttttga 3480
QY 3421 TGAATAAAATCATCTAACATGTGCCAAAACGCTTTTTCAGATTTCGAGATTTTTGA 3480
Db 3481 ttttttcaatcaagatatgcttaataacacatatataatcatcaatgtgaattcttg 3540
QY 3481 TTTTTCATTCAGATATGCTTATTAACACATATATAATCATTAATGTGAATTCTTG 3540
Db 3541 tagaaatttgggcttctgcttagtatgctctacttttgaattgtcgaagaaaaa 3600
QY 3541 TAGAAATTTGGGCTTTCGTCTAGTATGCTCTACTTTTGAATTTGCTCAACGAAAAA 3600
Db 3601 tcatgtggttgttcatatgatgaccaaataagcaattttatatatttccctat 3660
QY 3601 TCATGTGTTTGTTCATATGATGACGAAAAAATAGCAATTTTATATATTTCCTTAT 3660

CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.6%; Score 996; DB 9; Length 7653;

Best Local Similarity 99.8%; Pred. No. 0.00e+00;

Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Db 2701 ttgaagtgaatatatttttactgaagtcgagtgattatttttaacacta 2760
      |||||||
QY 2701 TTGAAGTGAATATATTTTACTGAAAGCTCGAGTGATTATTTTAAACACTA 2760
      |||||||

Db 2761 atttcgtgcgcaaaagccattttagatttgcgaaataacttgcacacacac 2820
      |||||||
QY 2761 ATTTTCGTGGCGCAAAAGCCATTGTGAGATTGCCGAAATACTCTGCACACACAC 2820
      |||||||

Db 2821 acacacatctccttcaaatatcccttttccagtggtgactcgaatgctgctgaattcga 2880
      |||||||
QY 2821 ACACACATCTCCTTCAATATCCCTTTTCCAGTGTGACTCGAATGCTGTGCAATTGCA 2880
      |||||||

Db 2881 gtgtccaatgtcacccgcaagccatcgtcggacgcgcatgtgaagcccgccgctaacac 2940
      |||||||
QY 2881 GTGTCCAATGTACCCGCAAGCCATCGTCGGAGCCGCCGATTGAGCCCGCGCTACAC 2940
      |||||||

Db 2941 ttaccgacccgagtcacccgtgacagcgtctcttcagtggtcatcatcattcattatca 3000
      |||||||
QY 2941 TTCACCGACCCGAGTTCACCGTGCAGCGCTCTTCAGTGTCATCATCTCTTATCA 3000
      |||||||

Db 3001 ggatatctactcaagagcaagatctcgttctcgatcgctgcactcatcatcgatcg 3060
      |||||||
QY 3001 GGATATCTACTCAAGAGCAAGATCTCGTCTCGATCGCGTGACATTTCATCGGATCG 3060
      |||||||

Db 3061 acacaattattcatctctccagtcacagccatttccagccaaccttgaattgtatgcg 3120
      |||||||
QY 3061 ACACAATTATTCTCTCCAGTCAACGCCATTCCAGCCAACTTGTATGTGATGCG 3120
      |||||||

Db 3121 aacactaaattctgagaatgcgcattactcaacataattgacgcgcaaatatctcgtagc 3180
      |||||||
QY 3121 AACACTAAATTCTGAGATGCGCATTAACAATATTGACGCGCAAAATATCTCGTAGC 3180
      |||||||

Db 3181 gaaaaatacagtaaccctttaaagtactatgtagtgatgattacggcctcgatttcg 3240
      |||||||
QY 3181 GAAAAATACAGTAACCTTTAAATGACTATTGTAGTGTGATTTACGGGCTCGATTTTCG 3240
      |||||||

Db 3241 aaacgaatatatgctcgaaatgtgacaacgaatttcaatttgcatttctgttctt 3300
      |||||||
QY 3241 AACGATATATGCTCGAATGTGACACGAATTTAATTTGTCAATTTTGTGTTTCTT 3300
      |||||||

Db 3301 ttgatatatttgcataaataatttccgttaaacagacacacgagtgatcagttact 3360
      |||||||
QY 3301 TTGATATTTTGATCAATTAATAATTATTTCCGTAACAGACACACGCGCTACAGTACT 3360
      |||||||

Db 3361 cttttaaagagtlacagtagtttgcgttcaagataatttgaagaattttaaacttt 3420
      |||||||
QY 3361 CTTTAAAGAGTTACAGTAGTTTTCGCTCAAGATATTTTGAAGAAGATTTTAAACATTT 3420
      |||||||

Db 3421 tgaaaaaaatcatctaacatgtgccaagccttttcaagttcgcagatttttga 3480
      |||||||
QY 3421 TGAATAAAATCATCTAACATGTGCCAAACGCTTTTTCAGATTTCGAGATTTTGTGA 3480
      |||||||

Db 3481 ttttttcatccaagatatgtcttatacacacataaattatcatcattgaatttcttg 3540
      |||||||
```

```
QY 3481 TTTTTCATCAAGATATGCTTATTACACATATAATTATCATTAATGTGAATTCTTG 3540
      |||||||
Db 3541 tagaaatttgggcttctcgcttctagtagtgccttactttagaattgtccaagaaaaa 3600
      |||||||
QY 3541 TAGAAATTTTGGGCTTTTCGTTCTAGTAGTCTTACTTTTGAATTTGCTCAACGAAAAA 3600
      |||||||

Db 3601 tcatgtgttgcataatgaatgaccaaataagcaattttatataatttccctat 3660
      |||||||
QY 3601 TCATGTGTTGTTCATATGATGACGAAAAATAGCAATTTTATATATTTTCCCTAT 3660
      |||||||

Db 3661 tcatgtgtgcagaaaaatagtaaaaaagcgatcattt 3700
      |||||||
QY 3661 TCATGTGTGACAGAAAAATAGTAAAAAAGCGCATGCAATTT 3700
      |||||||

RESULT 9
ID 064742 standard; DNA; 7653 BP.
AC 064742;
DT 23-JUN-1994 (first entry)
DE ced-3 (C6434T) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key
FH Location/Qualifiers
FT repeat_unit
  /tag= a
  /rpt_type= INVERTED
  /note= "Inverted w.r.t. repeat at 1490-1614"
  1490..1614
FT repeat_unit
  /tag= b
  /rpt_type= INVERTED
  /note= "Inverted w.r.t. repeat at 1356-1472"
  2232..2366
FT exon
  /tag= c
  /number= Exon_1
  2367..2429
FT intron
  /tag= d
  /number= Intron_1
  2450..2575
FT exon
  /tag= e
  /number= Exon_2
  2576..2853
FT intron
  /tag= f
  /number= Intron_2
  2854..3107
FT exon
  /tag= g
  /number= Exon_3
  3108..4302
FT intron
  /tag= h
  /number= Intron_3
  3126..3243
FT repeat_unit
  /tag= i
  /rpt_type= INVERTED
  /note= "Inverted w.r.t. repeat at 3329-3396"
  3329..3396
FT repeat_unit
  /tag= j
  /rpt_type= INVERTED
  /note= "Inverted w.r.t. repeat at 3126-3243"
  3487..3759
FT repeat_unit
  /tag= k
  /rpt_type= INVERTED
  /note= "Inverted w.r.t. repeat at 3782-4070"
  3782..4070
FT repeat_unit
  /tag= l
  /rpt_type= INVERTED
  /note= "Inverted w.r.t. repeat 3487-3759"
  4303..4634
FT exon
  /tag= m
  /number= Exon_4
  4635..5546
FT intron
  /tag= n
```



```

OY 3361 CTTTAAAGAGTTACAGTAGTTTTCGTTCAAGATATTTTGAAGAATTTTAAACAATT 3420
Db 3421 tgaataaaatcatcctaacaatgtgccaaaacgcttttccaagttcgagatttttga 3480
OY 3421 TGAATAAAATCATCTACATGTGCCAAAACGCTTTTTCAGTTTCGAGATTTTGA 3480
Db 3481 ttttttcatcagaatagtcttataacacacataatataatcaatgaattcttg 3540
OY 3481 TTTTTCATCAAGATATGCTTATTAACACATATAATATCATTAATGTGAATTCTTG 3540
Db 3541 tagaatttgggcttctcgttctagtatgctctactttgaaattgctcaacgaaaaa 3600
OY 3541 TAGAAATTTGGGCTTTTCGTTCTAGATGCTCTACTTTGAAATTGCTCAACGAAAAA 3600
Db 3601 tcatgtgttcttcataatgaatgacccaaaatagcaatttttatatatattccctat 3660
OY 3601 TCATGTGTTGTTCATATGAATGACGAAAAATAGCAATTTTATATATTTTCCCTAT 3660
Db 3661 tcatgtgtgcagaaaaaatagtaaaaaagcgcatgcat 3700
OY 3661 TCATGTGTGCAGAAAAATAGTAAAAAGCGCATGCAATT 3700

```

RESULT 8

```

ID Q64745 standard; DNA; 7653 BP.
AC Q64745;
DT 23-JUN-1994 (first entry)
DE ced-3 (C7020T) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT FT /*tag= a
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t repeat at 1490-1614"
FT FT 1490..1614
FT FT /*tag= b
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT FT 2232..2366
FT FT /*tag= c
FT FT /number= Exon_1
FT FT 2367..2429
FT FT /*tag= d
FT FT /number= Intron_1
FT FT 2450..2575
FT FT /*tag= e
FT FT /number= Exon_2
FT FT 2576..2853
FT FT /*tag= f
FT FT /number= Intron_2
FT FT 2854..3107
FT FT /*tag= g
FT FT /number= Exon_3
FT FT 3108..4302
FT FT /*tag= h
FT FT /number= Intron_3
FT FT 3126..3243
FT FT /*tag= i
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT FT 3329..3396
FT FT /*tag= j
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT FT 3487..3759
FT FT /*tag= k
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3782-4070"

```

```

FT repeat_unit 3782..4070
FT FT /*tag= l
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 3487-3759"
FT FT 4303..4634
FT FT /*tag= m
FT FT /number= Exon_4
FT FT 4635..5546
FT FT /*tag= n
FT FT /number= Intron_4
FT FT 4688..4719
FT FT /*tag= o
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5221..5330
FT FT /*tag= p
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5547..5760
FT FT /*tag= q
FT FT /number= Exon_5
FT FT 5761..5814
FT FT /*tag= r
FT FT /number= Intron_5
FT FT 5815..5942
FT FT /*tag= s
FT FT /number= Exon_6
FT FT 5943..6297
FT FT /*tag= t
FT FT /number= Intron_6
FT FT 6062..6138
FT FT /*tag= u
FT FT /rpt_type= INVERTED
FT FT 6298..6537
FT FT /*tag= v
FT FT /number= Exon_7
FT FT 6538..7012
FT FT /*tag= w
FT FT /number= Intron_7
FT FT 6567..6625
FT FT /*tag= x
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT FT 6905..6965
FT FT /*tag= y
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 6567-6625"
FT FT 7012..7075
FT FT /*tag= z
FT FT /number= Exon_8
FT FT 7020
FT FT /*tag= w
FT FT /note= "C>T, from allele n1163"
FT FT W09325685-A.
FT FT 23-DEC-1993.
FT FT 14-JUN-1993; U05701.
FT FT 12-JUN-1992; US-897788.
FT FT 20-NOV-1992; US-979638.
FT FT (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT FT Horvitz HR, Shaham S, Yuan J;
FT FT WPI; 94-007542/01.
FT FT P-PSDB; R53288.
FT FT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT FT develop agents to increase or prevent cell death in organisms
FT FT Claim 14; Fig 4; 127pp; English.
FT FT The sequences given in Q64735-45 represent mutations of the C. elegans
FT FT ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT FT was most abundant in embryos, but was also detected in larvae and young
FT FT adults, suggesting that ced-3 is expressed not only in cells undergoing
FT FT cell death. The four largest introns as well as sequences 5' of the
FT FT start codon contain repetitive elements, some of which have been
FT FT characterised in non-coding regions of other C. elegans genes, such
FT FT as fem-1, lin-12 and myod. The Ced-3 protein is 503 amino acids in

```


FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT repeat_unit
FT /tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT repeat_unit
FT /tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT repeat_unit
FT /tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT exon
FT /tag= m
FT /number= Exon_4
FT /tag= n
FT /number= Intron_4
FT repeat_unit
FT /tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT repeat_unit
FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon
FT /tag= q
FT /number= Exon_5
FT /tag= r
FT /number= Intron_5
FT intron
FT /tag= s
FT /number= Exon_6
FT /tag= t
FT /number= Intron_6
FT mutation
FT /tag= u
FT /note= "C>T, from allele n717"
FT repeat_region
FT /tag= v
FT /rpt_type= INVERTED
FT exon
FT /tag= w
FT /number= Exon_7
FT /tag= x
FT /number= Intron_7
FT repeat_unit
FT /tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit
FT /tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon
FT /tag= aa
FT /number= Exon_8
PN WO9325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI: 94-007542/01.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms

PS Claim 14; Fig 4; 127pp; English.
CC The sequences given in Q64735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

Query Match 99.6%; Score 996; DB 9; Length 7653;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 2701 ttgaagtgaatatatttttactgaagcgcgagtgattatttttaacacta 2760
QY 2701 TTGAAGTGAATATATTTTACTGAAGCTCGAGTGATTTATTTTAAACACTA 2760
Db 2761 atttcgtgcgcaaaagccattttagatttgcgaaataactgtcacacacac 2820
QY 2761 ATTTTCGTGGCGCAAAAGCCATTGTGATTTGCCGAATACTTGTCACACACAC 2820
Db 2821 acacacatctccttcaaatatcccttttccagtggtgactcgaatgctgtaattcga 2880
QY 2821 ACACACATCTCCTTCAATATCCCTTTTCCAGTGTGACTCGAATGCTGTGCAATTGCA 2880
Db 2881 gtgtccaatgtcacccgcaagccatcgtcgagccgcgcattgagcccgccgctaacac 2940
QY 2881 GTGTCCAATGTCACCCGCAAGCCATCGTCGAGCCGCGCATTGAGCCCGGCTAACAC 2940
Db 2941 ttcaaccgacccgagttcacggtgacagcgtctcttcagtgatcatcattcattca 3000
QY 2941 TTCACCGACCCGAGTTCACCGGTGACAGCGTCTCTTCAGTGCATCATCTTATCA 3000
Db 3001 ggatattctactcaagagaagatctcgttctcgatcgctgacattcatcattcgatcg 3060
QY 3001 GGATATTCTACTCAAGAGCAGATCTCGTTCGATCGCGTGACACTTCATTCATCGATCG 3060
Db 3061 acacattattcatctcctccagtcacgcatcttccagccaaacctgtatgtgatgag 3120
QY 3061 ACACATTATTCTATCTCTCCAGTCACGCAATTTCCAGCCAACTGTATGTGATGCG 3120
Db 3121 aacactaaattctgagaatgagcattactcaacatatgtgacgagcaaatatctcgtagc 3180
QY 3121 AACACTAAATTCTGAGAATGCGCATTTACTCAACATATTTGACGCGCAATATCTCGTAGC 3180
Db 3181 gaaaaatacagtaacccttaaatgactattgtagtgatgattacgggctcgatttcg 3240
QY 3181 GAAAAATACAGTAACCCCTTAAATGACTATTGTAGTGTGATTTACGGGCTCGATTTCG 3240
Db 3241 aaacgaatatatgctgaatgtgacaacgaattttaatttgtcatttttgttttctt 3300
QY 3241 AAACGAATATATGCTGAATGTGACAACGAATTTTATTTGTGTTTCTT 3300
Db 3301 ttgatatttttgatcaatataaatatttccgtaaacagacaccagcggtacagtaact 3360
QY 3301 TTGATATTTTGTGATCAATTAATAATTATTTCCGTAACAGACACAGCGGTACAGTACT 3360
Db 3361 ctttaagaggttacagtagtttcgcttcaagatatatttgaaagaattttaacattt 3420

PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.
DR P-PSDB; R53287.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127p; English.
CC The sequences given in Q64735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
CC sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

Query Match 99.6%; Score 996; DB 9; Length 7653;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 2701 ttgaagtgaatatatttattactgaaagctcgagtgattattatttttaacacta 2760
QY 2701 TTGAAGTGAATATATTATTACTGAAAGCTCGAGTGATTATTATTTTTAACACTA 2760
Db 2761 atttcgtggcgcaaaagccattttgtagattgcccgaataactgtcacacacacac 2820
QY 2761 ATTTTCGTGGCGCAAAAGGCCATTTGTAGATTGGCGAAATACTTGTACACACACAC 2820
Db 2821 acacacatcctcctaataatccctttccagtggtgactcgaatgctgtcgaaatcga 2880
QY 2821 ACACACATCTCCTTCAAAATATCCCTTTTCCAGTGTGACTCGAATGCTGCAATTGCA 2880
Db 2881 gtgtcaatgtcacccgcaagccatcgtcgagccgagcatgtgaccccgccgctacac 2940
QY 2881 GTGTCAATGTCAACCGGCAAGCCATCGTCGAGCCGGCATGTGAGCCCCGGCGTACAC 2940
Db 2941 ttcacgaccccgagttcacccgtgacagcgctccttcagtgatcatcattcatttataca 3000
QY 2941 TTCACGACCCGAGTTCACCCGTCACAGCGCTCTTCAGTGTCAATCATCTCTTATCA 3000
Db 3001 ggatatctactcaagagcaagatcgttctcgatcgcggtgcaacttcattcgcgatcg 3060
QY 3001 GGATATCTACTCAAGAGCAAGATCTGTTCTCGATCGCGTGCACCTTCATTCAATCGGATCG 3060
Db 3061 acacaattatcatctcctccagtcacagcatttcccagccaacctgtatgttatgagcg 3120
QY 3061 ACACAATTATTCATCTCCTCCAGTCAACGCATTTCGCCAGCCAACCTTGTATGTGATGCG 3120
Db 3121 aacactaaatctgagaatcgcatctactcaacataattgaacgagcaaatatcgttagc 3180
QY 3121 AACACTAAATCTGAGAATCGCAATTACTCAACATATTGACGCGCAAAATATCTCGTAGC 3180
Db 3181 gaaaataacagtaacccttaaatgactatgtagtgtgattacgggtcgatttcg 3240
QY 3181 GAAAAATACAGTAACCCCTTAATGACTATTGTAGTGTGCAATTACGGGCTCGATTTTCG 3240

Db 3241 aacgaatatatgctcgaattgtgacaagaaatttaattgttcattttgtgttctt 3300
QY 3241 AAACGAATATATGCTCGAATTGTGACAAACGAATTTAAATTGTGATTTTGTCTTCTT 3300
Db 3301 ttgatatlttgatcaatnaaattatttccgtaaacagacaccagcggtacagtact 3360
QY 3301 TTGATATTTTGTGATCAATTAATAATTATTTCGGTAACAGACACCAGCGCTACAGTACT 3360
Db 3361 cttaaagaagttacagtagtttgccttcaagataatttgaagaattttaaacttt 3420
QY 3361 CTTTAAAGAGTTTACAGTAGTTTTCGCTTCAAGATATTTTGAAAAGAAATTTAAACATTT 3420
Db 3421 tgaataaaatcatcctaactgtgccaaaagccttttcaagtttgcgagatttttga 3480
QY 3421 TGAATAAAATCATCTAACATGTGCCAAAACGCTTTTTCAAAGTTTCGAGATTTTGA 3480
Db 3481 ttttttcaatcaagatatgcttataacacataataattcaatgtgaattcttg 3540
QY 3481 TTTTTCATTCAGATATGCTTATTAACACATATAATTATCATTAATGTGAATTTCTTG 3540
Db 3541 tagaaatttgggctttcgtctagtatgctctactttgaaattgtcgaagaaaaa 3600
QY 3541 TAGAAATTTTGGGCTTTTCTCTAGTATGCTCTACTTTTGAATTTGCTCAACGAAAAA 3600
Db 3601 tcatgtgttgttcatatgaatgaccaaataagcaatttlatatatttccctat 3660
QY 3601 TCATGTGTTGTTCATATGATGACGAAAAATAGCAATTTTATATATTTCCTTAT 3660
Db 3661 tcatgtgttgcagaaaaatagtaaaaaagcgcatgcat 3700
QY 3661 TCATGTGTGCAGAAAAATAGTAAAAAAGCGCATGCAATT 3700

RESULT 7
ID Q64739 standard; DNA; 7653 BP.
AC Q64739;
DT 23-JUN-1994 (first entry)
DE ced-3 (G6297A) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key
FT repeat_unit
FT location/Qualifiers
FT 1356..1472
FT /*tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT 1490..1614
FT /*tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT 2232..2366
FT /*tag= c
FT /number= Exon_1
FT 2367..2429
FT /*tag= d
FT /number= Intron_1
FT 2450..2575
FT /*tag= e
FT /number= Exon_2
FT 2576..2853
FT /*tag= f
FT /number= Intron_2
FT 2854..3107
FT /*tag= g
FT /number= Exon_3
FT 3108..4302
FT /*tag= h
FT /number= Intron_3
FT 4316..4324
FT /*tag= i
FT repeat_unit


```
QY 3061 ACACAAATTATTCATCTCCTCCAGTCAACGCATTTCCAGCCAACTTGTATGTTGATCGG 3120
Db 3121 aacctaaattctgagaatgagcgaatctcaacatatctgacgcgaatatctcgtagc 3180
QV 3121 AACACTAAATCTTGAGATGCGGCTTACTCAACATATTTGACGCGCAATATCTCGTAGC 3180
Db 3181 gaaaaatacagtaaccctttaatgactatltgtagtgctgatttaacggctcgatttcg 3240
QY 3181 GAAAAATACAGTAACCCCTTAATAGCATATGTAGTGTGCAATTAACGGCTCGATTTTCG 3240
Db 3241 aaacgaatatatgctcgaattgtgacaacgaattttaattgtcatttttgtgtttctt 3300
QY 3241 AAACGAATATATGCTCGAATGTGACCAACGAATTTAATTTGTCAATTTTGTGTTCTT 3300
Db 3301 ttgatatatttgatcaatlaataattattccgtaaacagacacccagcggtacagttact 3360
QY 3301 TTGATATTTTGTGATCAATTAATAAATTAATTTCCGTAACAGACACACGCGCTACAGTACT 3360
Db 3361 ctttaagaggttacagtagtttcgcttcaagatatltgaaaaagaattttaacattt 3420
QY 3361 CTTTAAAGAGGTACAGTAGTTTCGCTTCAAGATATTTGAAAAAATTTTAAACATTT 3420
Db 3421 tgaaaaaaaatcatctaacatgtgccaacaaacgctttttcaagtttcgcagatttttga 3480
QY 3421 TGAATAAAATCATCTAACATGTGCCAAAACGCTTTTTCAGATTTTCGAGATTTTGTGA 3480
Db 3481 ttttttcatccaagatatgcttatttaacacatatattatcatttaattgtgaatttcttg 3540
QY 3481 TTTTTCATTCACAGATATGCTTATTAACACATATTAATTAATTAATGTAATTTCTTG 3540
Db 3541 tagaaatttgggcttctcgtctagtagtctcacttctgaattgtcacaagaaaaaa 3600
QY 3541 TAGAATTTTGGGCTTTTCGTTTGTAGTATGCTCTACTTTGAAATTTGCTCAACGAAAAA 3600
Db 3601 tcatgtggttctcatatgaatgacaaataatagcaattttatatatttccctat 3660
QY 3601 TCATGTGCTTGTTCATATGAATGACGAAAAATAGCAATTTTATATATTTTCCCTATP 3660
Db 3661 tcatgtgtgcagaaaaaatagtaaaaaaagcgcatgcat 3700
QY 3661 TCATGTGTGCGAGAAAAATAGTAAAAAAGCGCATGCAATTT 3700

RESULT 6
ID Q64744 standard; DNA; 7653 BP.
AC Q64744;
DT 23-JUN-1994 (first entry)
DE ced-3 (G6536A) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key
FT repeat_unit Location/Qualifiers
FT 1356..1472
FT /*tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT 1490..1614
FT /*tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT 2232..2366
FT /*tag= c
FT /number= Exon_1
FT 2367..2429
FT /*tag= d
FT /number= Intron_1
FT 2450..2575
FT /*tag= e
FT /number= Exon_2
FT 2576..2853
FT /*tag= f
FT intron
FT PN WO9325685-A.

/number= Intron_2
2854..3107
/*tag= g
/number= Exon_3
3108..4302
/*tag= h
/number= Intron_3
3126..3243
/*tag= i
/rpt_type= INVERTED
/note= "Inverted w.r.t. repeat at 3329-3396"
3329..3396
/*tag= j
/rpt_type= INVERTED
/note= "Inverted w.r.t. repeat at 3126-3243"
3487..3759
/*tag= k
/rpt_type= INVERTED
/note= "Inverted w.r.t. repeat at 3782-4070"
3782..4070
/*tag= l
/rpt_type= INVERTED
/note= "Inverted w.r.t. repeat 3487-3759"
4303..4634
/*tag= m
/number= Exon_4
4635..5546
/*tag= n
/number= Intron_4
4688..4719
/*tag= o
/rpt_type= INVERTED
/note= "Repeat 3"
5221..5330
/*tag= p
/rpt_type= INVERTED
/note= "Repeat 3"
5547..5760
/*tag= q
/number= Exon_5
5761..5814
/*tag= r
/number= Intron_5
5815..5942
/*tag= s
/number= Exon_6
5943..6297
/*tag= t
/number= Intron_6
6062..6138
/*tag= u
/rpt_type= INVERTED
6298..6537
/*tag= v
/number= Exon_7
6538
/*tag= w
/note= "G>A, from allele n2426"
6538..7012
/*tag= x
/number= Intron_7
6567..6625
/*tag= y
/rpt_type= INVERTED
/note= "Inverted w.r.t. repeat at 6905-6965"
6905..6965
/*tag= z
/rpt_type= INVERTED
/note= "Inverted w.r.t. repeat 6567-6625"
7012..7075
/*tag= aa
/number= Exon_8
```



```
FT /number= Exon_1
FT 2367..2429
FT /tag= d
FT /number= Intron_1
FT 2450..2575
FT /tag= e
FT /number= Exon_2
FT 2576..2853
FT /tag= f
FT /number= Intron_2
FT 2854..3107
FT /tag= g
FT /number= Exon_3
FT 3108..4302
FT /tag= h
FT /number= Intron_3
FT 3126..3243
FT /tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT 3329..3396
FT /tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT 3487..3759
FT /tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT 3782..4070
FT /tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT 4303..4634
FT /tag= m
FT /number= Exon_4
FT 4635..5546
FT /tag= n
FT /number= Intron_4
FT 4688..4719
FT /tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5221..5330
FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5547..5760
FT /tag= q
FT /number= Exon_5
FT 5761..5814
FT /tag= r
FT /number= Intron_5
FT 5815..5942
FT /tag= s
FT /number= Exon_6
FT 5943..6297
FT /tag= t
FT /number= Intron_6
FT 6062..6138
FT /tag= u
FT /rpt_type= INVERTED
FT 6298..6537
FT /tag= v
FT /number= Exon_7
FT 6372
FT /tag= w
FT /note= "G>A, from allele n1286"
FT 6538..7012
FT /tag= x
FT /number= Intron_7
FT 6567..6625
FT /tag= y
FT /rpt_type= INVERTED
```

```
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit
FT 6905..6965
FT /tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT 7012..7075
FT /tag= aa
FT /number= Exon_8
FT
FT WO9325685-A.
FT 23-DEC-1993.
FT 14-JUN-1993; U05701.
FT 12-JUN-1992; US-897788.
FT 20-NOV-1992; US-979638.
FT (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT Horvitz HR, Shaham S, Yuan J;
FT WPI; 94-007542/01.
FT P-PSDB; R53284.
FT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT develop agents to increase or prevent cell death in organisms
FT PS Claim 14; Fig 4; 127p; English.
FT The sequences given in Q64735-45 represent mutations of the C. elegans
FT CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT CC was most abundant in embryos, but was also detected in larvae and young
FT CC adults, suggesting that ced-3 is expressed not only in cells undergoing
FT CC cell death. The four largest introns as well as sequences 5' of the
FT CC start codon contain repetitive elements, some of which have been
FT CC characterised in non-coding regions of other C. elegans genes, such
FT CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
FT CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
FT CC region that might be a transmembrane region. One region of Ced-3 is
FT CC very rich in serine. It is thought that this region is involved in
FT CC protein-protein interactions, similar to acid blobs in transcription
FT CC factors. Of the mutations which occur within the ced-3 gene, eight of
FT CC the mutations are missense mutations, two are nonsense mutations and
FT CC two are putative splicing mutations. These mutations establish the
FT CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
FT CC function is not essential for viability. The ced-3 and ced-4 gene
FT CC products may be used to develop agents for treating conditions
FT CC characterised by cell deaths, such as myocardial infarction, stroke,
FT CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
FT CC infection, aging or hair loss.
FT SQ Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

Query Match 99.6%; Score 996; DB 9; Length 7653;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 2701 ttgaagtgaatatatttattactgaagctcgagtgattatttttaacta 2760
QY 2701 TTGAAGTGAATATATTATTACTGAAGCTCGAGTGATTATTTTTAACCTA 2760
Db 2761 atttcgtgcgcaaaagccattgtagatttgcgcaaatctgtcacacacac 2820
QY 2761 ATTTGCGTGGCGCAAAAGCCATTGTGATTGCCGAAATACCTGTGCACACACAC 2820
Db 2821 acacacatcctccaatatccctttccagtggtgactcgaatgctgcaatcga 2880
QY 2821 ACACACATCTCCTTCAATATCCCTTTTCCAGTGTGACTCGAATGCTGCAATCGA 2880
Db 2881 gtgtccaatgtcacccgcaagccatcgtgcgagccgcgcatgagcccccgcgtacac 2940
QY 2881 GTGTCCAATGTACACCGCAAGCCATCGTCGAGCCGCCGCAATGAGCCCCGCCGCTACAC 2940
Db 2941 ttaccgcgaccgagttcacccgtgacagcgtctcttcaagtgtcatcatcacttctatca 3000
QY 2941 TTCACCGACCCGAGTTCACCCGTGACAGCGTCTCTTCAGTGTCATTCATTCCTTATCA 3000
Db 3001 ggatatctactcaagagcaagatctcgttctcgatcgcggtcgaacttcatcgcgacg 3060
QY 3001 GGATATCTACTCAAGAGCAAGATCTCGTTCGATCGCGGTGCACCTTCATTCATCGGATCG 3060
Db 3061 acacaattatcatctcctccagtcacgcatttccacgcaacccctgtatgtgacg 3120
|||||
```



```

FT      exon      6298..6537
FT      /tag= w
FT      /number= Exon_7
FT      intron    6538..7012
FT      /tag= x
FT      /number= Intron_7
FT      repeat_unit 6567..6625
FT      /tag= y
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT      repeat_unit 6905..6965
FT      /tag= z
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      exon      7012..7075
FT      /tag= aa
FT      /number= Exon_8
FT      /tag= aa
FT      /number= Exon_8
PD      .WO9325685-A.
PD      23-DEC-1993.
PF      14-JUN-1993; U05701.
PR      12-JUN-1992; US-897788.
PR      20-NOV-1992; US-979638.
PA      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI      Horvitz HR, Shaham S, Yuan J;
PI      WPI; 94-007542/01.
DR      P-PSDB; R53282.
PT      Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT      develop agents to increase or prevent cell death in organisms
PS      Claim 14; Fig 4; 127pp; English.
CC      The sequences given in Q64735-45 represent mutations of the C. elegans
CC      ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC      was most abundant in embryos, but was also detected in larvae and young
CC      adults, suggesting that ced-3 is expressed not only in cells undergoing
CC      cell death. The four largest introns as well as sequences 5' of the
CC      start codon contain repetitive elements, some of which have been
CC      characterised in non-coding regions of other C. elegans genes, such
CC      as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC      length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC      region that might be a transmembrane region. One region of Ced-3 is
CC      very rich in serine. It is thought that this region is involved in
CC      protein-protein interactions, similar to acid blobs in transcription
CC      factors. Of the mutations which occur within the ced-3 gene, eight of
CC      the mutations are missense mutations, two are nonsense mutations and
CC      two are putative splicing mutations. These mutations establish the
CC      null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC      function is not essential for viability. The ced-3 and ced-4 gene
CC      products may be used to develop agents for treating conditions
CC      characterised by cell deaths, such as myocardial infarction, stroke,
CC      degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC      infection, aging or hair loss.
SQ      Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

```

```

Query Match      99.6%; Score 996; DB 9; Length 7653;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2941      |||||
QY      TTCACCGACCCGAGTTACCCGTGACAGCGTCTCTTCCAGTGTCAATCAATCACTTCTATCA
Db      3001      ggatatctactcaagagcaagatctcgtctcgtatcgcgctgacacttcattcatcgatcg
QY      3001      GGATATCTACTCAAGAGCAAGATCTCGTTCTCGATCGCGTGACATTCATTCGATCG
Db      3061      acacaattatcattcctccocagtcacagcatttccocagccaacctgtatgtatgacg
QY      3061      ACACATATATCATCTCTCCAGTCAACGCATTTCCAGCCACACCTTGATGTGATGCG
Db      3121      aacactaaattctgagaatgacgcatctactcaacatatctgacgacgcaaatatctcgtacg
QY      3121      AACACTAAATTCGAGAAATGCGCATTTACTCAACATATTTGACCGCCCAATATCTCGTAGC
Db      3181      gaaaaatacagtaacccctttaaatagtactgtagtgctgattacgagcctcgatttcg
QY      3181      GAAAAATACAGTAACCCCTTAAATGACTAATGTAGTGTGATTAACGGGCTCGATTTTCG
Db      3241      aaacgaatatatgctcgaattgtgacaagcaatttaattgtcattttgtgttctc
QY      3241      AAACGAATATATGCTCGAATTTGTGACAAAGCAATTTTAATTTGTGATTTTGTGTTCTT
Db      3301      ttgatatttttgatcaatlaataattttccgtaaacagacacaccagcggtacagctact
QY      3301      TTGATATTTTGTGATCAATTAATAATTAATTTCCGTAACAGACACACCGCTACAGTACT
Db      3361      ctttaagagttacagtagtttctcgtctcaagatatcttgaaaaagaattttaaacttc
QY      3361      CTTTAAAGAGTTACAGTAGTTTTCGCTTCAAGATATTTTGAAGAATTTTAAACATTT
Db      3421      tgaaaaaaatcatcctaacaatggtgccaacacgcttttccaagtttcgcagatttctga
QY      3421      TGAATAAAATCATCTAATCATGTGCCAAACGCTTTTTCAGGTTTCGAGATTTTGA
Db      3481      tttttcatccaagatatgcttatttaacacatatattatcatcaatgtgaattctgtg
QY      3481      TTTTTCATTCAGATATGCTTATTAAACATATTAATTATCATTAATGGAATTTCTTG
Db      3541      tagaaattttggccttctcgtctctagtagtctacttttgaattgtcacaagaaaaa
QY      3541      TAGAATTTTGGGCTTTTCGTTCTAGTAGCTCTACTTTGAATTCCTCAACGAAAAA
Db      3601      tcatgtgtttgttcatatgaaatgaccacaaaatagcaatttttatatatcttccctat
QY      3601      TCATGTGTTTGTTCATATGAATGACGAAAAATAGCAATTTTATATATTTTCCCTAT
Db      3661      tcatgtgtgacgaaaaaataagtaaaaaagcgcatgcatltt
QY      3661      TCATGTTGTGCAAGAAAAATAGTAAAAAAGCGCATGCAATTT 3700

```

```

RESULT 5
ID      Q64741 standard; DNA; 7653 BP.
AC      Q64741;
DT      23-JUN-1994 (first entry)
DE      ced-3 (G6372A) gene.
KW      C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW      embryogenesis; cell death; hydrophilic; transmembrane; region;
KW      hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW      protein synthesis; null phenotype; calcium-binding domain; ss.
OS      Caenorhabditis elegans.
FH      Key
FT      repeat_unit 1356..1472
FT      /tag= a
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 1490-1614"
FT      repeat_unit 1490..1614
FT      /tag= b
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 1356-1472"
FT      exon      2232..2366
FT      /tag= c

```


Db 2821 acacacatctccttcaaatatcccttttccagtggtgactcgaatgctgtcgaattcga 2880
|||||
QY 2821 ACACACATCTCCTTCAAAATATCCCTTTTCCAGTGTGACTCGAATGCTGCGAATTGCA 2880
Db 2881 gtgtccaatgtcaccggcgaagccatcgtcggagccgcgcgcatltgagcccccgcgcgtacac 2940
|||||
QY 2881 GTGTCCAATGTCCACGGCAAGCCATCGTCGGAGCCGCCGCAATTGAGCCCCCGCGCTACAC 2940
Db 2941 ttcaccgaccccgagttcacccgtgacacgcgtctcttcagltgcatcatcttcttataca 3000
|||||
QY 2941 TTCACCGACCCGAGTTCAACCGTGACAGCGCTCTTCAAGTGCATCATTCACCTTCTTATCA 3000
Db 3001 ggatactactcaagagcaagatcgtctcgcgacgcgtgcacttcattcgcgacgcg 3060
|||||
QY 3001 GGATATCTACTCAAGACCAAGATCTGTTCTCGATCGCGTGCACCTTCATTCATCGGATCG 3060
Db 3061 acacaattatctatctcctccagtcacgcgcatctccagccaacctgtatgttgatgcg 3120
|||||
QY 3061 ACACAATTATCTACTCTCCTCCAGTCAACGCATTTCCAGCCAACCTGTATGTTGATGCG 3120
Db 3121 aacactaaattctgagaatgcgcattactcaacataattgacgcgcaaatatctcgtagc 3180
|||||
QY 3121 AACACTAAATTCAGGAATGCGCATTTACTCAACATATTTGACGCGCAATATCTCGTAGC 3180
Db 3181 gaaaaatacagtaaccctttaaatagtactatgtgagtgctgatttaacggcgcgatttcg 3240
|||||
QY 3181 GAAAAATACAGTAACCTTTAAATGACTATGTAGTGTGATTTACGGGCTCGATTTTTCG 3240
Db 3241 aaacgaatatagtctcgaattgtgacaaacgaatttaattgtcatlttctgttcttct 3300
|||||
QY 3241 AAACGAATATATGCTCGAATTGTGACAAACGAATTTAAATTGTGATTTTGTGTTTCTT 3300
Db 3301 ttgataatttgcataaataaataattatctccgtaaacagacacacgcggtacagttact 3360
|||||
QY 3301 TTGATATTTTGTGATTAATTAATTAATTTCCGTAACAGACACACGCGCTACAGTACT 3360
Db 3361 ctttaagagttacagtagtttctcgcctcaagataattgaaaagaattttaaacattt 3420
|||||
QY 3361 CTTTAAAGAGTACAGTAGTCTTTCGCTCAAGATATTGAAAAGAATTTTAAACATTT 3420
Db 3421 tgaaaaaaaatcatcctaacaatgtgccaacgaagcttttttcaagtttcgcagatttttga 3480
|||||
QY 3421 TGAATAAAAAATCATCTAACATGTGCCAAAAACGCTTTTTCAGATTTCGACGATTTTGA 3480
Db 3481 ttttttcatccaagataatgcttataaacacataaattatcatlaatgtgaatttcttg 3540
|||||
QY 3481 TTTTTCATTCATCAAGATATGCTTATTAACACATATAATTAATCATTAATGTGAATTTCTTG 3540
Db 3541 tagaaatttgggcttttcgltcgtatgctactacttgaattgtcgaacgaaaaaa 3600
|||||
QY 3541 TAGAAATTTTGGGCTTTTCTGTTCTAGTATGCTACTTTTGAATTTGCTCAACGAAAAAA 3600
Db 3601 tcatgtggttgcataatgaatgaccaaataagcaatttttatataattttccctat 3660
|||||
QY 3601 TCATGTGTTTGTTCATATGAATGACGAAAAAATAGCAATTTTATATATATTTCCCTAT 3660
Db 3661 tcatgttgcagaaaaaatagtaaaaaagcgcatgcatltt 3700
|||||
QY 3661 TCATGTGTGTCAGAAAAATAGTAAAAAAGCGCATGCAATTT 3700

RESULT 4
ID Q64738 standard; DNA: 7653 BP.
AC Q64738;
DT 23-JUN-1994 (first entry)
DE ced-3 (C5940T) gene.
KW C. elegans; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH key Location/Qualifiers
FT repeat_unit 1356..1472

FT /tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t repeat at 1490-1614"
FT 1490..1614
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT 2232..2366
FT /tag= c
FT /number= Exon_1
FT 2367..2429
FT /tag= d
FT /number= Intron_1
FT 2450..2575
FT /tag= e
FT /number= Exon_2
FT 2576..2853
FT /tag= f
FT /number= Intron_2
FT 2854..3107
FT /tag= g
FT /number= Exon_3
FT 3108..4302
FT /tag= h
FT /number= Intron_3
FT 3126..3243
FT /tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT 3329..3396
FT /tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT 3487..3759
FT /tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT 3782..4070
FT /tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT 4303..4634
FT /tag= m
FT /number= Exon_4
FT 4635..5546
FT /tag= n
FT /number= Intron_4
FT 4688..4719
FT /tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5221..5330
FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5547..5760
FT /tag= q
FT /number= Exon_5
FT 5761..5814
FT /tag= r
FT /number= Intron_5
FT 5815..5942
FT /tag= s
FT /number= Exon_6
FT 5943..6297
FT /tag= t
FT /note= "C>T, from allele n1165"
FT 5943..6297
FT /tag= u
FT /number= Intron_6
FT 6062..6138
FT /tag= v
FT /rpt_type= INVERTED


```
NT 23-JUN-1994 (first entry)
DE ced-3 (G2487A) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key location/Qualifiers
FT repeat_unit 1356..1472
FT FT /tag= a
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t repeat at 1490-1614"
FT FT 1490..1614
FT FT /tag= b
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT FT 2232..2366
FT FT /tag= c
FT FT /number= Exon_1
FT FT 2367..2429
FT FT /tag= d
FT FT /number= Intron_1
FT FT 2450..2575
FT FT /tag= e
FT FT /number= Exon_2
FT FT 2487
FT FT /tag= f
FT FT /note= "G>A, from allele n718"
FT FT 2576..2853
FT FT /tag= g
FT FT /number= Intron_2
FT FT 2854..3107
FT FT /tag= h
FT FT /number= Exon_3
FT FT 3108..4302
FT FT /tag= i
FT FT /number= Intron_3
FT FT 3126..3243
FT FT /tag= j
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT FT 3329..3396
FT FT /tag= k
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT FT 3487..3759
FT FT /tag= l
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT FT 3782..4070
FT FT /tag= m
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 3487-3759"
FT FT 4303..4634
FT FT /tag= n
FT FT /number= Exon_4
FT FT 4635..5546
FT FT /tag= o
FT FT /number= Intron_4
FT FT 4688..4719
FT FT /tag= p
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5221..5330
FT FT /tag= q
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5547..5760
FT FT /tag= r
FT FT /number= Exon_5
FT FT 5761..5814
FT FT /tag= s
FT FT /number= Intron_5
```

```
FT exon 5815..5942
FT FT /tag= t
FT FT /number= Exon_6
FT FT 5943..6297
FT FT /tag= u
FT FT /number= Intron_6
FT FT 6062..6138
FT FT /tag= v
FT FT /rpt_type= INVERTED
FT FT 6298..6537
FT FT /tag= w
FT FT /number= Exon_7
FT FT 6538..7012
FT FT /tag= x
FT FT /number= Intron_7
FT FT 6567..6625
FT FT /tag= y
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT FT 6905..6965
FT FT /tag= z
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 6567-6625"
FT FT 7012..7075
FT FT /tag= aa
FT FT /number= Exon_8
FT FT
FT PN W09325685-A.
FT PD 23-DEC-1993.
FT PF 14-JUN-1993; U05701.
FT PR 12-JUN-1992; US-897788.
FT PR 20-NOV-1992; US-979638.
FT PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT PI Horvitz HR, Shaham S, Yuan J;
FT DR WPI; 94-007542/01.
FT DR P-PSDB; R53280.
FT PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
FT PT develop agents to increase or prevent cell death in organisms
FT PS Claim 14; Fig 4; 127pp; English.
FT CC The sequences given in Q64735-45 represent mutations of the C. elegans
FT CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
FT CC was most abundant in embryos, but was also detected in larvae and young
FT CC adults, suggesting that ced-3 is expressed not only in cells undergoing
FT CC cell death. The four largest introns as well as sequences 5' of the
FT CC start codon contain repetitive elements, some of which have been
FT CC characterised in non-coding regions of other C. elegans genes, such
FT CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
FT CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
FT CC region that might be a transmembrane region. One region of Ced-3 is
FT CC very rich in serine. It is thought that this region is involved in
FT CC protein-protein interactions, similar to acid blobs in transcription
FT CC factors. Of the mutations which occur within the ced-3 gene, eight of
FT CC the mutations are missense mutations, two are nonsense mutations and
FT CC two are putative splicing mutations. These mutations establish the
FT CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
FT CC function is not essential for viability. The ced-3 and ced-4 gene
FT CC products may be used to develop agents for treating conditions
FT CC characterised by cell deaths, such as myocardial infarction, stroke,
FT CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
FT CC infection, aging or hair loss.
FT SQ Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

Query Match 99.6%; Score 996; DB 9; Length 7653;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 998; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 2701 ttgaagtgaatatatttactgaagctcgagtgattatttactttaacacta 2760
Qy 2701 TTGAAGTGAATATATTATTACTGAAGCTCGAGTGATPATTTATTTTAACACTA 2760
Db 2761 atttcgtgycgaagccatttgtagatttcgaaataactgtcacacacacac 2820
Qy 2761 ATTTGCGTGGCGCAAAAGCCATTTGTAGATTTCGCAAAATACCTGTACACACACAC 2820
```


FT intron 5943..6297
FT /tag= x
FT /label= Intron-6
FT repeat_unit 6062..6138
FT /tag= y
FT /label= Repeat-4
FT exon 6298..6537
FT /tag= z
FT allele 6322
FT /tag= aa
FT /label= T(n1949)
FT /note= "creates premature stop codon at 412"
FT allele 6372
FT /tag= ab
FT /label= A(n1286)
FT allele 6434
FT /tag= ac
FT /label= T(n1129,n1164)
FT /note= "causes A449V mutation"
FT allele 6485
FT /tag= ad
FT /label= T(n2430)
FT /note= "causes A466V mutation"
FT allele 6535
FT /tag= ae
FT /label= A(n2426)
FT /note= "causes E483K mutation"
FT intron 6538..7012
FT /tag= af
FT /label= Intron-7
FT repeat_unit 6567..6625
FT /tag= ag
FT /label= Repeat-5
FT repeat_unit 6905..6965
FT /tag= ah
FT /label= Repeat-5
FT exon 7013..7653
FT /tag= ai
FT /note= "in-frame stop codon at 7073-7075"
FT allele 7020
FT /tag= aj
FT /label= T(n1163)
FT /note= "causes S486F mutation"
PN W09625946-A1.
PD 29-AUG-1996.
PF 23-FEB-1996; U02473.
PR 24-FEB-1995; US-394189.
PI (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 96-425082/42.
DR P-PSDB; R98754.
PT Ced-3 and human interleukin 1-beta convertase genes and proteins
PT useful to treat inflammation and diseases characterised by cell
PT death
PS Claim 18; Fig 3; 139pp; English.
CC The Caenorhabditis elegans ced-3 gene (T38196) was cloned by
CC mapping RFLPs and chromosome walking, and genomic DNA cloned
CC in plasmid pJ107 was sequenced. EMS-induced alleles were also
CC sequenced. The gene codes for a cell death protein (R98754) that
CC is structurally similar to human interleukin-1 beta converting
CC enzyme (ICE) (R98755), suggesting that Ced-3 protein may be a
CC cysteine protease like ICE and that ICE may be a human equivalent
CC of the nematode cell death gene. The ced-3 gene can be used as a
CC probe or in the prodn. of Ced-3 protein and novel drugs for
CC enhancing or inhibiting the activity of ICE, ced-3 and related
CC genes for the treatment of inflammatory diseases and/or diseases
CC caused by cell death. Novel inhibitors of ced-3 activity include
CC portions of the ced-3 gene and its product.
SQ Sequence 7653 BP; 2429 A; 1456 C; 1270 G; 2498 T;

Query Match 100.0%; Score 1000; DB 24; Length 7653;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2701 ttgaagtgaaatatatttacttgaaagctcgagtgatattatttttaacacta 2760
QY 2701 TTGAAGTGAATATATTATTACTGAAAGCTCGAGTGATATTATTTTAACACTA 2760
Db 2761 atttcgtgcgcaaaagccattttgtgatttgccgaaatattgtcacacacac 2820
QY 2761 ATTTTCGTGGCGCAAAAGCCATTTTGTGATTTGCCGAAATATTGTCACACACAC 2820
Db 2821 acacacatctctccaatatatcccttttccagtggttactcgaaatgtcgtcgaattcga 2880
QY 2821 ACACACATCTCCTCAATATATCCCTTTTCCAGTGTGACTCGAATGCTGCGAATTGCA 2880
Db 2881 gtgtccaatgtcaccgcaagccatcgtcgagcgcgcatlgaagcccgccggtacac 2940
QY 2881 GTGTCCAATGTCAACCGCAAGCCATCGTCGGAAGCCGCGCAITGAGCCCGCGGTACAC 2940
Db 2941 ttcaaccgaccgagttcaaccgtgacagcgctcttcaagtgtcatcattcacttctatca 3000
QY 2941 TTCACCGACCCGAGTTCAACCGTGAACGCGTCTTCACTGTCAATTCACCTTCTATCA 3000
Db 3001 ggatatctactcaagagcaagatctgcttcgtatcgatcgcggtgacattcatcgatcg 3060
QY 3001 GGATATCTACTCAAGACCAAGATCTCGTCTCGATCGCGTGCATTCATTCGATCG 3060
Db 3061 acacaattatcatctcctccagtgcaacgcatlccccaagccacttgtatgtgatcg 3120
QY 3061 ACACAATTATTCATCTCCTCCAGTCAACGCATTTCCAGCCCAACCTTGATGTGATGCG 3120
Db 3121 aacactaaattctgagaatgagcattgacattcaacatatltagccgcgcaaatatctcgtagc 3180
QY 3121 AACACTAAATTCTGAGAATGCCATTACTCAACATATTGACGCGCAATATCTCGTAGC 3180
Db 3181 gaaaaatacacagtaacccctttaaattgactatltgagtgctgacttcaagcgctcgatttcg 3240
QY 3181 GAAAAATACAGTAACCCCTTAAATGACTATTGTAGTGTGATTTACGGGCTCGATTTCG 3240
Db 3241 aaacgaatatagctcgaattgtgacaacgaatttaattgtgcatlgtgttctt 3300
QY 3241 AAACGAATATATGCTCGAATTGTGACAACGAATTTAATTGTGATTTGTGTTCTT 3300
Db 3301 ttgatatttttgatcaattaataattatcttcgtaaacagacaccagcgctacagtaact 3360
QY 3301 TTGATATTTTGTGATCAATTAATAATTATTTCGTAACACAGACACACGCGTACAGTACT 3360
Db 3361 cttttaagagttacagtagtttctcgcttcaagatatltagaaagaattttaacattt 3420
QY 3361 CTTTAAAGAGTTACAGTAGTTTCCGTTCAAGATATTTTGAAGAAATTTTAAACATTT 3420
Db 3421 tgaaaaaaaatcatctaacatgtgcccacaaagcctttttcaaglttcgcagatttttga 3480
QY 3421 TGAATAAAATCATCTAATCATGTGCCAAACGCTTTTTCAGATTTCGAGATTTTTGA 3480
Db 3481 ttttttcatccaagatatgcttatttaacacatatataattcatltaattgtgaattttg 3540
QY 3481 TTTTTCATTTCAAGATATGCTTATTAAACACATATATTATTAATGTAATTTCTTG 3540
Db 3541 tagaaatttgggcttctcgcttagtatgctctactttgaaattgtccaagaaaaa 3600
QY 3541 TAGAAATTTGGGCTTTTCGTTAGTATGCTTACTTTGAAATTGCTCAACGAAAAA 3600
Db 3601 tcatgtgtgttgcataatgaatgaacgaacaaatagcaatttttatatatattccctat 3660
QY 3601 TCATGTGTGTGTTCATATGAATGACGAAAAATAGCAATTTTATATATTTCCCTAT 3660
Db 3661 tcatgtgtgtgcagaaaataagtaaaaaaagcgcatgcat 3700
QY 3661 TCATGTGTGTGCAGAAAAATAGTAAAAAAGCGCATGCATTT 3700

RESULT 3
ID Q64736 standard; DNA; 7653 BP.
AC Q64736;


```
QY 2701 TTGAAGTGAATATATTTTACTGAAGCTCGAGTGATTTATTTTAACTA 2760
Db 2761 atttcgtgcgcaaaagccattttagatttgcgaaaaataacttgcacacacacac 2820
QY 2761 ATTTCTGGCGCAAAAGCCATTGTAGATTGCCGAAATACTTGTACACACACAC 2820
Db 2821 acacacatctccttcaaatatcccttttccagtggtgactcgaaatgctgcgaatcga 2880
QY 2821 ACACACATCTCTCTCAAAATATCCCTTTTCCAGTGTGACTCGAATGCTGCGAATTGCA 2880
Db 2881 gtgtccaatgtcacccgcaagccatcgtagagccgcatgagcccgccgctacac 2940
QY 2881 GTGTCCAATGTCAACCGCAAGCCATCGTAGCCGCGCATTTGAGCCCGCGCTACAC 2940
Db 2941 ttcacccgaccgagttcacccgtgacagcgctctctcagtgatcattcacttctatca 3000
QY 2941 TTCACCGACCCGAGTTCAACCGTAGACGCTCTCTCAGTGCATCATTCACCTCTTATCA 3000
Db 3001 ggatatctactcaagagcaagatctcgttctcgatcgctgcacttcattcatcgatcg 3060
QY 3001 GGATATCTACTCAAGAGCAAGATCTCTCTCGATCGCGTGCACCTCATTCATCGGATCG 3060
Db 3061 acacaattatcatctcctccagtcacagcatttcccagccaaacctgtatgttgatcg 3120
QY 3061 ACACAATATTCATCTCTCTCCAGTCAACGCATTTCCAGCCCAACCTTGATGTGATGCG 3120
Db 3121 aacacataatctgagaatgagcattactcaacataattgacgagcaaatatctcgtagc 3180
QY 3121 AACACTAAATTTCTGAGAATGGCATTACTCAACATATTTGACGCGCAAAATATCTCGTAGC 3180
Db 3181 gaaaaatacagtaacccttaaatgactattgtagtgtcgaattacgggctcgatttcg 3240
QY 3181 GAAAAATACAGTAACCCCTTAAATGACTATTGTAGTGTGATTTACGGGCTCGATTTCG 3240
Db 3241 aaacgaatatatctcgaattgtgacacgaatttlaatttgcatttltgttctt 3300
QY 3241 AAACGAATATATGCTCGAATTGTGACACGAATTTTAAATTTGTCATTTTGTGTTTCTT 3300
Db 3301 ttgatatatttgatcaatlaataatatttccgtaaacagacacagcgctacagtagt 3360
QY 3301 TTGATATTTTTCATCATTAATAATTTTCCGTAACAGACACACGCGCTACAGTACT 3360
Db 3361 cttttaaagagttacagtagtlttcgcttcaagatatatttgaagaatttlaaacatt 3420
QY 3361 CTTTAAAGAGTTACAGTAGTTTCCGTTCAAGATATTTTGAAGAATTTTAAACATTT 3420
Db 3421 tgaaaaaaaatcatctaacatgctgccaaaacgcttttccaagtttcgcagatttttga 3480
QY 3421 TGAATAAAATCATCTAATGTCGCAAAACGCTTTTTCAGATTTCGAGATTTTTGA 3480
Db 3481 ttttttcatcaagatatgcttataacacataataattatcalttaagtgaattcttg 3540
QY 3481 TTTTTCATTCATCAAGATATGCTTATTAACACATATTAATTATTAATGTGAATTTCCTTG 3540
Db 3541 tagaaatttgggcttctcgttagtgcctactttgaaatgctcaacgaaaaaa 3600
QY 3541 TAGAAATTTTGGGCTTTCGTTCTAGTATGCTCTACTTTGAAATGCTCAACGAAAAAA 3600
Db 3601 tcatgtgttgcataatgaatgacgaaaaatagcaatttttatatatatttccctat 3660
QY 3601 TCATGTGTTGTTCATATGAATGACGAATAATAGCAATTTTATATATTTTCCCTAT 3660
Db 3661 tcatgtgtgcagaaaaaatagtaaaaaagcgcatgtcatt 3700
QY 3661 TCATGTGTGCAGAAAAATAGTAAAAAAGCGCATGCAATTT 3700
```

RESULT 2
ID T38196 standard; DNA; 7653 BP.
AC T38196;
DT 17-DEC-1996 (first entry)
DE Nematode Ced-3 gene.

```
KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;  
KW cell death; apoptosis; neural degeneration; inflammation;  
KW antiinflammatory; ds.  
OS Caenorhabditis elegans.  
FH Key location/Qualifiers  
FT repeat_unit 1356..1472  
FT FT /tag= a  
FT FT /label= Repeat-1  
FT FT 1490..1614  
FT FT /tag= b  
FT FT /label= Repeat-2  
FT FT 2167..2366  
FT FT /tag= c  
FT FT /codon_start= 2232..2234  
FT FT 2310  
FT FT /tag= d  
FT FT /label= T(n1040)  
FT FT /note= "causes L27F mutation"  
FT FT 2367..2429  
FT FT /tag= e  
FT FT /label= Intron-1  
FT FT 2430..2575  
FT FT /tag= f  
FT FT 2487  
FT FT /tag= g  
FT FT /label= A(n118)  
FT FT /note= "causes G65R mutation"  
FT FT 2576..2853  
FT FT /tag= h  
FT FT /label= Intron-2  
FT FT 2854..3107  
FT FT /tag= i  
FT FT 3108..4302  
FT FT /tag= j  
FT FT /label= Intron-3  
FT FT 3126..3243  
FT FT /tag= k  
FT FT /label= Repeat-1  
FT FT 3329..3396  
FT FT /tag= l  
FT FT /label= Repeat-1  
FT FT 3487..3759  
FT FT /tag= m  
FT FT /label= Repeat-2  
FT FT 3782..4070  
FT FT /tag= n  
FT FT /label= Repeat-2  
FT FT 4303..4633  
FT FT /tag= o  
FT FT 4634..5546  
FT FT /tag= p  
FT FT /label= Intron-4  
FT FT 4688..4719  
FT FT /tag= q  
FT FT /label= Repeat-3  
FT FT 5221..5330  
FT FT /tag= r  
FT FT /label= Repeat-3  
FT FT 5546..5760  
FT FT /tag= s  
FT FT 5757  
FT FT /tag= t  
FT FT /label= A(n2433)  
FT FT /note= "causes G360S mutation"  
FT FT 5761..5814  
FT FT /tag= u  
FT FT /label= Intron-5  
FT FT 5815..5942  
FT FT /tag= v  
FT FT 5940  
FT FT /tag= w  
FT FT /label= T(n1165)  
FT FT /note= "creates premature stop codon at 403"
```



```
FT      /tag= g
FT      /note= "G>A, fromm allele n718"
FT      2576..2853
FT      /tag= h
FT      /number= Intron_2
FT      2854..3107
FT      /tag= i
FT      /number= Exon_3
FT      3108..4302
FT      /tag= j
FT      /number= Intron_3
FT      3126..3243
FT      /tag= k
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3329-3396"
FT      3329..3396
FT      /tag= l
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3126-3243"
FT      3487..3759
FT      /tag= m
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3782-4070"
FT      3782..4070
FT      /tag= n
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 3487-3759"
FT      4303..4634
FT      /tag= o
FT      /number= Exon_4
FT      4635..5546
FT      /tag= p
FT      /number= Intron_4
FT      4688..4719
FT      /tag= q
FT      /rpt_type= INVERTED
FT      /note= "Repeat 3"
FT      5221..5330
FT      /tag= r
FT      /rpt_type= INVERTED
FT      /note= "Repeat 3"
FT      5547..5760
FT      /tag= s
FT      /number= Exon_5
FT      5757
FT      /tag= t
FT      /note= "G>A, from allele n2433"
FT      5761..5814
FT      /tag= u
FT      /number= Intron_5
FT      5815..5942
FT      /tag= v
FT      /number= Exon_6
FT      5940
FT      /tag= w
FT      /note= "C>T, from allele n1165"
FT      5943..6297
FT      /tag= x
FT      /number= Intron_6
FT      6062..6138
FT      /tag= y
FT      /rpt_type= INVERTED
FT      6297
FT      /tag= z
FT      /note= "G>A, fromm allele n717"
FT      6298..6537
FT      /tag= aa
FT      /number= Exon_7
FT      6322
FT      /tag= ab
FT      /note= "C>T, from allele n1949"
FT      6372
FT      /tag= ac
```

```
FT      /note= "G>A, from allele n1286"
FT      6434
FT      /tag= ad
FT      /note= "C>T, fromm alleles n1129 and n1164"
FT      6485
FT      /tag= ae
FT      /note= "C>T, from allele n2430"
FT      6535
FT      /tag= af
FT      /note= "G>A, from allele n2426"
FT      6538..7012
FT      /tag= ag
FT      /number= Intron_7
FT      6567..6625
FT      /tag= ah
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT      6905..6965
FT      /tag= ai
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      7013..7075
FT      /tag= aj
FT      /number= Exon_8
FT      7020
FT      /tag= ak
FT      /note= "C>T, from allele n1163"
FT      1..5850
FT      /tag= al
FT      /note= "Claim 2, inhibitor fragment"
FT      1..3020
FT      /tag= am
FT      /note= "Claim 2, inhibiotr fragment"
FT      WO9325694-A.
FT      23-DEC-1993.
FT      14-JUN-1993; U05705.
FT      12-JUN-1992; US-897788.
FT      20-NOV-1992; US-984182.
FT      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
FT      Horvitz HR, Shaham S, Yuan J;
FT      P-PSDB; R45262.
FT      WPI; 94-007551/01.
FT      develop drugs for treating conditions characterised by cell death
FT      or proliferation
FT      PS Disclosure; Fig 3; 132pp; English.
FT      This sequence encodes the C.elegans cell death gene, ced-3. Fragments
FT      of the amino terminal of the protein encoded by this sequence act as
FT      inhibitors of ced-3. This gene has considerable similarity to human
FT      interleukin-1beta convertase (ICE), which converts pro-interleukin-
FT      1beta to the active cytokine and is involved in inflammatory response
FT      in humans. The similarity between the two sequences suggests that
FT      inhibitors of ced-3 may also act as inhibitors of ICE. Human ICE and
FT      nematode Ced-3 proteins have an overall amino acid similarity of 28%.
FT      The ced-3 inhibitors may be used for identifying agents which affect
FT      the activity of a gene belonging to the to the ced-3/ICE family of
FT      genes and for diagnosis of diseases characterised by cell death. They
FT      can also be used to develop drugs for treating conditions characterised
FT      by cell deaths such as myocardial infarction, stroke, degenerative
FT      disease, traumatic brain injury, hypoxia, pathogenic infection, or
FT      hair loss, or drugs for reducing the proliferative capacity or size
FT      of a population of cells such as cancerous cells, cells which produce
FT      autoreactive antibodies, infected cells, hair follicle cells or cells
FT      which are critical to the life of a parasite, pest or recombinant
FT      organism. They may also be used in the diagnosis of inflammatory
FT      disease.
FT      CC Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;
FT      SQ
FT      Query Match 100.0%; Score 1000; DB 9; Length 7653;
FT      Best Local Similarity 100.0%; Pred. No. 0.00e+00;
FT      Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Db 2701 ttgaagtgaatatatttactgaagctcgagtgattattttaaacta 2760


```

/db_xref="PID:e1344529"
/db_xref="PID:g3874792"
/db_xref="GI:3874792"
/db_xref="SPTREMBL:P90772"
/translation="MDVODIIGIALILTAGSLAIIALPGCGGKNAKSAEGESSLS
SKKDGAKTPDGREGSEAGTSPSPGSAAPSPAPGCTHDPNYQTLAGIGNDCEYKKDGA
PAAGCGRAPANAKPGMAATHDPNYQTLAGLNNVFEKKDGAAPAGCGGPAAPADO
NAKAATHDPNYQTLAGLNNVFEKKDGGGAPATPGGKAPANIGAKAATHDPNYQTL
AGLNNVFEKKDGGGAGGAPAGAPKPLQPMNKNOKAATLDPNYQTLAAVGDVFGADK
KVGAGAPCGKKPIQPNNKOGAKAATMDPNYQTLAAVGDVFGADKKKK"

```

Query Match	11.2%;	Score 112;	DB 21;	Length 39211;
Best Local Similarity	80.6%;	Pred. No. 4.38e-39;		
Matches 233;	Conservative	0;	Mismatches 43;	Indels 13; Gaps 4;
Db	8014	GAACACAAATCTGAGAAATGCCGTAATTGCACACATATTTGACGCGCAAAATATCTCGTA	8073	
Qy	3120	GAACACTAAATCTGAGAAATGCCATTACTCAACATATTTGACGCGCAAA-TATCTCGTA	3178	
Db	8074	GCGAAAACTACAGTAATTCCTTAAATGGCTACTGTAGCGCTTATTTGATTCACGGGCTC	8133	
Qy	3179	GCGAAAAATACAGTAACCCCTTTAAATGACTATTGTAGTG-----TCGATTTACGGGCTC	3232	
Db	8134	GATTTTGAATGAATTTCTTTTCGAATAGTGACAACGA---TAATTTGTAATTTTAA	8189	
Qy	3233	GATTTTCGAACGAATATATGCTCGAATTTGTGACAACGAATTTTAAATTTGTCAATTTTGT	3292	
Db	8190	GTTTTTTTCTAATATTTT--ATCGAATTAATAATGATGTCCGTAATCGACGCCAACGCT	8247	
Qy	3293	GTTTTCTTTGATATTTTTTGATCAATTAATAAATATTTCCGTAACAGACACACGCGCT	3352	
Db	8248	ACAGTACTCACCGAATGAATTAATCTGTAGTTTTCGCTACGAGATATTTG	8296	
Qy	3353	ACAGTACTCTTTTAAAGAGTTACAGTAGTTTTCGCTCAAGATATTTTG	3401	

Search completed: Fri Aug 6 04:46:36 1999
Job time : 1857 secs.


```

elegans cDNA yk402c6.5; coded for by C. elegans cDNA
yk483d4.5; coded for by C. elegans cDNA yk431h12.5; coded
for by C. elegans cDNA yk273a10.5; coded for by C. elegans
cDNA yk222h8.5"
/codon_start=1
/protein_id="AAC78491.1"
/db_xref="PID:g3893872"
/translation="MSLARQNHSEVEAAVNKOINIELYASYVYLMSFYFDRDDVAL
PNIAKFEKQSDDEEREHATELMRVQNLGRGRVYLQIDKPENDEWGTALKAFELAL
EKFNESLKLKHLSTAGNHNDAMLDTFIEKYLDEQVKSINEFARMVANLKRVGSGVE
YVFDEKHEFSD"
complement(22450. .23347)
/gene="D1037.4"
complement(join(22450. .22583,22654. .22948,23008. .23068,
23116. .23270,23318. .23347))
/gene="D1037.4"
/notes="contains similarity to the RAS family (Pfam:
PF00071, score=220.5, E=2.4e-62, N=1); coded for by C.
elegans cDNA yk480a12.3; coded for by C. elegans cDNA
yk451h3.3; coded for by C. elegans cDNA yk282h8.3; coded
for by C. elegans cDNA yk404e3.3; coded for by C. elegans
cDNA yk246g5.3; coded for by C. elegans cDNA yk439c11.3;
coded for by C. elegans cDNA yk204a8.5; coded for by C.
elegans cDNA yk439c11.5; coded for by C. elegans cDNA
yk451h3.5; coded for by C. elegans cDNA yk480a12.5; coded
for by C. elegans cDNA yk404e3.5; coded for by C. elegans
cDNA yk246g5.5; coded for by C. elegans cDNA yk282h8.5"
/codon_start=1
/protein_id="AAC78494.1"
/db_xref="PID:g3893875"
/db_xref="GI:3893875"
/translation="MAKTYDYLFLKLLIGDSGVGKTCVLFRESDSFNNSFISTIGID
FKIRTIELDGKRIKLIOWDTAGQERFTITTAAYRGAMGILVYDITNERSFENIKMW
ARINIEEHAASDVERMIIGNKCDIEERREYSDRGEOLAIETGTFLETSAKANLIDE
AFETIARDIKSKMEQNMRAATGAALCFKFIALSPCSALFEVDCDSIQFNFSVPLYF
SYFLSV"
24090. .29335
/gene="D1037.1"
join(24090. .24176,24252. .24535,24591. .25001,25078. .25364
25794. .25942,26565. .26801,26960. .27050,27104. .27315,
27741. .27863,28804. .29032,29085. .29335)
/gene="D1037.1"
/notes="contains similarity to a double-stranded RNA
binding motif (Pfam: PF00035, score=42.7, E=8.1e-09, N=1);
coded for by C. elegans cDNA yk213d8.5; coded for by C.
elegans cDNA yk356a7.5; coded for by C. elegans cDNA
yk19c11.5; coded for by C. elegans cDNA yk476g4.5; coded
for by C. elegans cDNA yk205c1.5; coded for by C. elegans
cDNA yk208b5.5; coded for by C. elegans cDNA yk246d9.5;
coded for by C. elegans cDNA yk19c11.3; coded for by C.
elegans cDNA yk213d8.3; coded for by C. elegans cDNA
yk401c9.3; coded for by C. elegans cDNA yk208b5.3; coded
for by C. elegans cDNA yk476g4.3; coded for by C. elegans
cDNA yk356a7.3; coded for by C. elegans cDNA yk246d9.3;
coded for by C. elegans cDNA yk176c4.3"
/codon_start=1
/protein_id="AAC78493.1"
/db_xref="PID:g3893874"
/db_xref="GI:3893874"
/translation="MVDGSGITPQTCSEDIKDLINDVKSSEKPFVSVDKLAGCDQTFD
VGRPKKHHRHKSCKSKAYRRRLSKTWSSSRISRRSRSGRSQERRRSKRSKRSR
SLRKHREKIEKKLLEKAPKRLILPOAPSTENGSKPEKPEPRKDEADDLPGADFS
TPLYKETKLAISITTVLPRAENKKEVEDEIEELIAYGDGDFKEPEGIEKLENKNICE
NGNDNSENEKSEKDKNRKKEKTERKKRSHSSSDBRGRTSKESRRSRSDTRT
RRRSERSSSTETRNERPEYRWRWSRERKEYSRRRRRSRSREVIDKEKLLAIA
KTKAEAMSNGDRDSASIEDYVTCCKLQRRQERREKRRAGHDVSDHSDIDRETVRYKH
PYAMEKEPIRINIVATASANAQAKITGPEEPTQLGAQLRIVFPVSSGAVHKEAEM
VPEKEDVPKSCSVEQKKHALTSLEIDRCRAQGIAPAPAPVPKFLNSILPPAPPO
FLNPPTDSFRPTPPLPSVIYVKNKDOMILKEPKDIPAPSDIGKVLQRTQAQR
VLSTPNDFDAHARLANEQLISMAALKSLPGEYTGIGLRLITADLEQPQNKYHAW
VKKDFENAPATGGIGRLMEKMGWRGEGELGATGNLEPLMDVKSDRKGLTAE
EMSIQQRNKAQNNVAVDLSNKNPISLIMELCAKRRNPPSPSCSESGADHLKMFVW

```

gene	TIVINDVEYRPMCGSKQKKEGKAVALQVALLGLVLEMDNLPVYM"
CDS	complement(30411..33302) /gene="D1037.5" complement(join(30411..30756,30801..31122,31662..31948,

```

/3000="contains similarity to Ca2+-independent
phospholipases; coded for by C. elegans cDNA yk231g7.5"
/codon_start=1
/protein_id="AAC78490.1"
/db_xref="PID:g3893871"
/db_xref="GI:3893871"
/translation="MPREKSLSLNGKKVMFRYSKSNDRGKRQNSKDRARADHMAFDY

```

...
Note: remainder of annotations omitted.

Query Match	11.2%;	Score 112;	DB 22;	Length 36404;
Best Local Similarity	88.6%;	Pred. No. 4.38e-39;		
Matches 140;	Conservative	0;	Mismatches 16;	Indels 2;
			Gaps	2;

Db 21632 CTTAATTCGAGAATGCGTATTGCCAACATATTGACCGCGCAAATATCTCGTAGCGAA 21691

3125 CTAATTCGAGATGGCATTACTCACATATTTGACCGCAAA-TATCTCGTAGCGAA 3183

Db 21692 AACTACAGTAATCTTTAAATGACTACTGTAG-GTCGATTACGGGCTCTATTTTCGAA 21750

3184 AAATACAGTAACCTTTAAATGACTATTGTAGTGTCCGATTACGGGCTCGATTTTCGAAA 3243

Db 21751 CAAATTTTGTGTCGAATGTCACACGATATTC AATT 21788

QY 3244 CGAATATATGCTCGAATTGTGACCAACGAATTTAATT 3281

	CEC36F7	39211 bp	DNA	INV	23-NOV-1998
LOCUS					

ACCESSION 281045

VERSION 281045.1 GI:1627644

RNA; tRNA-Ile.

ORGANISM *Caenorhabditis elegans*
Phylum: Nematoda: Secernentea: Rhabditia: Rhabditida

Rhabditina; Rhabditoidea; Peloderinae; Caenorhabdit

AUTHORS
Lightning, J.
List of Submissions

JOURNAL
Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail:
[redacted]

REFERENCE	2 (bases 1 to 39211)
WILSON P. AINSCOURTH R. ANDERSON, K. BAYNES, C. BERKS, M.:	

Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson A., Crayton M., Dear, S., Du, Z., Durbin, B., Favello, A.,

Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,

Latreille, P., Lightning, J., Lloyd, C., McMurtry, A., Mollimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifken, H., Roopra, A.,

Saunders, D., Shownkeen, R., Smaldon, N., Smith, A., Sonnenhaumer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,

Vaughan, K., Waterston, K., Watson, A., Weinstock, L., Wilkinson-Sproat, J., and Wohlman, P.

TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome 11 of *A. elegans*

JOURNAL
MEDI. INF 94150718
Nature 368 (6466), 32-38 (1994)

COMMENT Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other

For a graphical representation of this sequence and its analysis available information.

see: <http://webace.sanger.ac.uk/cgi->

```
bin/displayrad=wordaccrclass-sequence 500000 0001/
```


/protein_id="AAC48078.1"
/db_xref="PID:g1397325"
/db_xref="GI:1397325"

Note: remainder of annotations omitted.

Query Match 12.4%; Score 124; DB 22; Length 38112;
Best Local Similarity 72.8%; Pred. No. 2.59e-45;
Matches 206; Conservative 0; Mismatches 76; Indels 1; Gaps 1;

Db 20832 GAACACTAATTTCTGAGATGCGATTTGCACACACACACTTGCAGCGCAAAATATCTCGTA 20891
|||||
QY 3120 GAACACTAATTTCTGAGATGCGATTTGCACACACACTTGCAGCGCAAAATATCTCGTA 3178
|||||

Db 20892 GCGAAACTACAGTATTTCTTAATGACTACTGTAGCGGTTGTGCGATTACGGGCTC 20951
|||||
QY 3179 GCGAAACTACAGTATTTCTTAATGACTACTGTAGCGGTTGTGCGATTACGGGCTGATTTT 3238
|||||

Db 20952 GTTTTTCGAAATTAATTTTCTGCTTTTTCGAAATGTTTTCGTAATTTATGTTT 21011
|||||
QY 3239 CGAAACGATATATGCTCGAATTTGACACACGAAATTTAATTTGTCATTTTGTGTTTC 3298
|||||

Db 21012 CTGTAAATAATTTACCGATTAATAATTTATTTCTGTAATCGACACACAGCGCTACAGTA 21071
|||||
QY 3299 TTTGATATTTTGTGATCAATTAATAATTTATTTCCGTAACACAGACACAGCGCTACAGTA 3358
|||||

Db 21072 CTATTTTAAAGAAATTTACTGTTCTCGCTACGTAATTTTG 21114
|||||
QY 3359 CTCTTTTAAAGAGTTACAGTAGTTTTCGCTCAAGATATTTTG 3401
|||||

RESULT 14
LOCUS CELD1037 36404 bp DNA INV 20-NOV-1998
DEFINITION Caenorhabditis elegans cosmid D1037.
ACCESSION AF106592
NID 93893870
VERSION AF106592.1 GI:3893870
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
AUTHORS
1 (bases 1 to 36404)
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Lloyd, C., McMurray, A., Mortimore, B.,
Latreille, P., Lightning, J., Percy, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Rife, L., Rife, L., Roopra, A.,
Saunders, D., Showkeen, R., Smaison, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierri-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Spratt, J., and Wolldman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL
MEDLINE
Nature 368 (6466), 32-38 (1994)
94150718
REFERENCE
AUTHORS
2 (bases 1 to 36404)
Ledwith, J. and Biewald, T.
The sequence of C. elegans cosmid D1037
JOURNAL
TITL
Unpublished (1998)
REFERENCE
AUTHORS
3 (bases 1 to 36404)
Waterston, R.
Direct Submission
Submitted (15-NOV-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England

e-mail: rwenematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' clone is T28F2, 200 bp overlap; 3' clone is H26D21, 3001 bp
overlap. Actual start of this clone is at base position 197 of
CELD1037; actual end is at 36404 of CELD1037

NOTES:

Coding sequences below are predicted from computer analysis, using
the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES

source

1. 36404
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"

/clone="D1037"
/chromosome="I"
complement(6412..11154)

/gene="D1037.2"
complement(6412..11154)

complement(join(6412..6613,7123..7204,7254..7358,
7484..7694,7750..8027,8346..8653,8764..8867,9218..9483,
9621..9741,9791..9978,10989..11154))

/gene="D1037.2"

/note="contains similarity to human putative DNA binding
protein SON3 (GB:X63753)"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC78492.1"

/db_xref="PID:g3893873"

/db_xref="GI:3893873"

/translation="MHPTTISEFLPLVQICRIYHCPNRHDKLSPTTPPLVFAF
NQSLSVSLFTVSRLSYGANELVILHKLPGELYLLGGRNHYNISISSMLEIARYE
WTSDEARNTCAAIISQTPASCENEFIRTYFELTNDLILGTHALOPTCAERKNAP
QLRISAVGMSPIDADSTAFIRSNENITVNAEISSEPLVLRNVIKMGCIENDV
ILRTPRGLSFEQANFLSMHKVKVGVVEVLFFSESPEMETEGGGLKVARVGRVCE
DDEGRSLYNKEWSSYEKARIECSIEETDTDFYNOFAGVAESPTSFYAFRSOLAG
IGASAIKSKSVISGSFASGKESTPDTCSRANDIEELSRIRLKLKIKISANPIY
IFHGKDFEVHVAQEDTRDLNRAYDILYVATNGLIKIVPSTDKTGRAVTLKVL
PTNSKIVDMSLSKNNLELIVITEQSVLKVPTATCHLASNCAECLAHPCHCAWADG
VDCIDIRTDORKSASODSGTCDVSEHNKPMIAPLKNKESAKIPVCLCETEKOKP
CATEVIOKEIVLTGSEFWKYLIFAVGVLTGSIHFYCFYFSRHIGDRASSSSRSSS
SMTRPITTTSSSARLPIDAFITSSIADEMFSSISANRPNFVSMSHSIRTYC"
complement(19578..20187)
/gene="D1037.3"

complement(join(19578..19679,19729..20082,20131..20187))
/gene="D1037.3"

/note="contains similarity to ferritins (Pfam: PF00210,
Score=311.3, E=8.8e-92, N=1); coded for by C. elegans CDNA
YK402c6.3; coded for by C. elegans CDNA YK333c6.3; coded
for by C. elegans CDNA YK447f6.3; coded for by C. elegans
CDNA YK483d4.3; coded for by C. elegans CDNA YK273a10.3;
coded for by C. elegans CDNA YK384g11.3; coded for by C.
elegans CDNA YK201e1.3; coded for by C. elegans CDNA
YK222h8.3; coded for by C. elegans CDNA YK162g4.3; coded
for by C. elegans CDNA YK365a7.3; coded for by C. elegans
CDNA YK68f7.5; coded for by C. elegans CDNA YK172d9.5;
coded for by C. elegans CDNA YK280h4.5; coded for by C.

NEIGHBORING COSMID INFORMATION:

The 3' cosmid is M01H9, 200 bp overlap. Actual start of this cosmid is at base position 1 of CELR08C7; actual end is at 38112 of CELR08C7

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES
source

location/Qualifiers

1. 38112

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="IV"

/clone="R08C7"

1277. 5453

/gene="R08C7.6"

join(1277. 1396,1737. 1890,2268. 2785,3313. 3850,

5065. 5335,5381. 5453)

/gene="R08C7.6"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC48075.1"

/db_xref="PID:g1397321"

/db_xref="GI:1397321"

/translation="MRLLLVFLVGVVVGONSQPTSGTPLDEGLKACEDAIIG
DICTITFRVLTEDDANEFCRLYAPWRLRAIRNEDGIPOTICEVEATMACDDGWI
OMFGYCFRHPYKNIQAEAIKVCNSLDGTEIAFLHRYIVGWKNYERGIQIM
VNGSETWQYVSTGAVDGLALALATGTVVESVPSNLIRIDPILRLGILCQYKE
RTLAIAVLGRYSEIYHQISIFLEDGVLIRSAHSYTRSKTNAELCKTLRPFVDVE
PEIPNDIMSALSDOYTEPWAISNSGAFETMDVSKRSSDVCLEKTHKYSVAPDPKM
PPEVLVETSVCGDNMSVAIVHSEIDTEPRMSDSQSVPMCKFGKRMNFVDLPKG
YTHYERANGMHVGHKYSMILPDKAMKYCDDGGAQLGGFNDKAEFEAIDKLISKV
HHWIGNRPRDCPGFTFHSDPNHQCGHNKIHEMLNVAATEDIPLEMGFKQPDHAWKP
QGRQEGSIAYGMPGNGEPKLDGSCCEVKMPFICTAKGVLKATATASRL"

complement(6651. 8720)

/gene="R08C7.7"

complement(join(6651. 6760,6809. 6986,7091. 7228,

7623. 7764,7867. 8168,8520. 8720))

/gene="R08C7.7"

/note="similar to C. elegans olfactory receptor Odr-10
(GB:U49449)"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC48081.1"

/db_xref="PID:g2911717"

/db_xref="GI:2911717"

/translation="MHLSEFNNIFLKIYQYTSAFMSVSLNCILIFLLIASHPKLGM
KYLMIYISSFEALYSIMDYTEPMVHSYKAFFVFRNFKNDSDFREHSFLIIVYCGC
FGFSIAIEGVHVVYRGAYVKEERDKMLGKRIYILFEMPIFYGTWMSVLCYLFHD
DSTDDYMRKTIEDGYDKIEDTSIIVLFHPVDKNGTSHDPDAVFASIAMFMILSS
VFSYFEGIKCYFRITEALSRCTCNVSVTKGLQQLFQALVQTFIPLIMYIPAIL
FTFPMIAVDIGFASSEVAMTIAVPAIDPLPMMFIKRYKAVFAFFSAIFRNPCCSI
RRAOCTQIELVSKSETPKNIGL"

8632. 11285

/gene="R08C7.5"

join(8632. 8864,9185. 9226,9273. 9384,9432. 9581,

9635. 9757,9980. 10558,10623. 10730,10777. 10902,

10949. 11164,11214. 11285)

/gene="R08C7.5"

/note="similar to other C. elegans chemoreceptors"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC48071.1"

/db_xref="PID:g1397317"

/db_xref="GI:1397317"

/translation="MRIQFNETDINAEVYCTIFRKIFONSRLCIVASTATDGLTIFIV
VAVGWAESHEIVEHFVDEDEMGHGRGKNGKNNKFNQFHRISQFETIKTVFTRSS
SRNHTKSDVAGAGPTWDEFLKVDGMSQKLQECENQFEKASNKVFAELFAKLVAKKV
ATEDFPAEQLOKLIDRNPCKPFGKDVKEIEAEHDELLETAMYLSDASEGGEITM

8632. 11285

/gene="R08C7.5"

join(8632. 8864,9185. 9226,9273. 9384,9432. 9581,

9635. 9757,9980. 10558,10623. 10730,10777. 10902,

10949. 11164,11214. 11285)

/gene="R08C7.5"

/note="similar to other C. elegans chemoreceptors"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC48071.1"

/db_xref="PID:g1397317"

/db_xref="GI:1397317"

/translation="MRIQFNETDINAEVYCTIFRKIFONSRLCIVASTATDGLTIFIV
VAVGWAESHEIVEHFVDEDEMGHGRGKNGKNNKFNQFHRISQFETIKTVFTRSS
SRNHTKSDVAGAGPTWDEFLKVDGMSQKLQECENQFEKASNKVFAELFAKLVAKKV
ATEDFPAEQLOKLIDRNPCKPFGKDVKEIEAEHDELLETAMYLSDASEGGEITM

8632. 11285

/gene="R08C7.5"

join(8632. 8864,9185. 9226,9273. 9384,9432. 9581,

9635. 9757,9980. 10558,10623. 10730,10777. 10902,

10949. 11164,11214. 11285)

/gene="R08C7.5"

/note="similar to other C. elegans chemoreceptors"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC48071.1"

/db_xref="PID:g1397317"

/db_xref="GI:1397317"

/translation="MRIQFNETDINAEVYCTIFRKIFONSRLCIVASTATDGLTIFIV
VAVGWAESHEIVEHFVDEDEMGHGRGKNGKNNKFNQFHRISQFETIKTVFTRSS
SRNHTKSDVAGAGPTWDEFLKVDGMSQKLQECENQFEKASNKVFAELFAKLVAKKV
ATEDFPAEQLOKLIDRNPCKPFGKDVKEIEAEHDELLETAMYLSDASEGGEITM

gene

CDS

NTPEPGEVYKAEELKAEHHDHMDTAMLLDPDSTODGEVYTMKPVSDISELSEED
PDETLIQNTLHMDNLSMKEELNADTIVTSTPIRACSTKIMVVPVREISPEVEKDIS
SSTHCRLNLNLPKMPVFTSMFGKKOEKPGVESHVSNLQKILNHRKRGKEVQKVS
SPPKLPKTVQPSKEEPTAKTAPPVKLLTSSPPKLPKTVQPKPKKPMESVK
TAAPPAKEIKPIVAKITEKAAEPAPVKEVPAKAKATAPKAPVMSYKMTPTRVHPRS
ATDYGESESDSEFERDKSKOKNFPKASDMNIVNEEYAKNLEFNVTFFGSMEKANL
TEIFGEMVKKKKRGSSKVMN"

complement(12024. 13414)

/gene="R08C7.8"

complement(join(12024. 12127,12172. 12310,12364. 12549,

12595. 12804,12969. 13196,13301. 13414))

/gene="R08C7.8"

/note="Similar to serine/threonine protein phosphatase"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC48076.1"

/db_xref="PID:g1397323"

/db_xref="GI:1397323"

/translation="MYEGESVVOAKRMIEHLKMGVTDFAFDKOIYVTLLEKAEETLK
PLPAMLQVEHPITIVGDHGLDALIRYDAVGYPKVOFLFDGDYVDRGAKSFEVSL
LLFCYKIRYPHLVFNELPLCARVGRILCMHGGISQNCNSWESKALKPNTPLTCD
EGLQVDLMADPTQDKCNTFAMNKQRAISVVGQKGLDDELTKLGISLIVRAHEVSOE
GFNLFNKKCIVYFSAPYCCGNDTNCGAIMHVSDSYELSTFLVRPMIATPENFEIVK
LMENNYKGLMVASPDPNRGRHLQSAVVPQPTPPSVQLVPPVTKVEIKQ"

14065. 16237

/gene="R08C7.4"

join(14065. 14201,14674. 14864,14909. 15006,15412. 15515,

15559. 15705,15783. 15954,16019. 16237)

/gene="R08C7.4"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC48080.1"

/db_xref="PID:g1397327"

/db_xref="GI:1397327"

/translation="MAKVRRKFFIVSFLLSILLFLQHSRQVKYIERTFILENGSO
NKRATCHYDCNRVEIHAARFKSRVPLAVITCPRRFGAEVSVSEDEEPOEPIP
LTVRAYDRPLHLSVGVPMYGEESKWEILEYVEHYRLVECHYRSKFSKVVNMID
DERLVYTGPLNLSYLRRLPNNGEVSTFNVLKTEPVPVKYSSDAQLMEDMLFKY
NKTIEISGOCCEGAEKIWACQVNTFFCNDREIYRFIDQCFKICIDFLNRYFLFGK
AIEFSKTLPKHGQKSEWELDGKLQDRTSRNPDIQVIREKINFGSETTGQICIRSAN
GEVRRNSRVVEFKVQEATAGL"

complement(17104. 20058)

/gene="R08C7.9"

complement(join(17104. 17259,17313. 18033,18907. 19806,

19982. 20058))

/gene="R08C7.9"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC48077.1"

/db_xref="PID:g1397324"

/db_xref="GI:1397324"

/translation="MAVFTLLNLEPEKSLNVLRRMPLTELIGFALISKTTKQAEERLN
VKMSLNSLDGAIRIHISDGSNDVPPPWFEKFLSVKPKQSKREFILMTSGGRNW
TNPERGVROFLDHVLEIMHLTKIYLCLEIDCDANTIRDMFDGLEISLQIFERNLVN
PFVQKMLTFSRQSEQVSWRGNPPLSDSHKIQAILGQFIHSVYLPISYFNDIIPNDL
LIDNVISICIGELRLRDVNIYLYKWIHGSNTKLECVRLYLDRVMRENITIEVLKNV
EYQMAPPEREFVIRAVGSEHEIRGGIIVQRNDSERIPSVFVKFLVSAQAPNDGKREF
RTKQAEERLNVMKMSVDVWIDEAIIKFIWADSDERIPSVFVKFLVSAQAPNDGKREF
LRTREGKTWPKFGARQFLDHALEVCNEVCIIKITLDCIDCDGNKICDMFDGLTIES
LSVCFEHLQLVNPFQQLTKFSRQSKMSVMGNPFPDDSHRIQIILCOFIHSVHL
AKLFSNVRLDGLLADAVLNMMDKMLKDVNIFLKHWHIGSNTKLESASLRLDAQ
MAPPEREFVIRANGSKEHIRGEIIVRRKDGTMITIVLNHENSYFKMFVHL"

complement(21564. 24678)

/gene="R08C7.10"

complement(join(21564. 21638,21690. 22022,22355. 22480,

22530. 22604,22659. 22890,22939. 23015,23065. 23153,

23205. 23674,23720. 23913,23960. 24114,24164. 24356,

24402. 24582,24638. 24678))

/gene="R08C7.10"

/note="similar to drosophila parallel sister chromatid
protein (GB:U40214)"

/codon_start=1

/evidence=not_experimental

BASE COUNT 9566 a 5644 c 5839 g 9473 t
ORIGIN

Query Match 13.1%; Score 131; DB 21; Length 30522;
Best Local Similarity 81.2%; Pred. No. 5.67e-49;
Matches 229; Conservative 0; Mismatches 44; Indels 9; Gaps 5;

Db 12857 AACACTAAATTCGAGATGCGTTTGTGCAACACATTTGACCGCGCAATATCTCGTAGC 12916
|||||
Qy 3121 AACACTAAATTCGAGATGCGCGATTACTACATATTTGACCGCGCAATATCTCGTAGC 3180
Db 12917 GAAACTACAGTAATTCGTTAAATGAGTAGCGCTGTGTCGATTACGGCTTAAATTC 12976
|||||
Qy 3181 GAAATAATACAGTAACCTTTAAATGACTATTG-TAGTGTGATTTACGGGCTCGATTTC 3239
Db 12977 GAAACGAATTCGG-TAGTTCCTTTTGTGTAAC--TTT--TTTGTCAATTTTCATGTTTCT 13030
|||||
Qy 3240 GAAACGAATATATGCTCGAATTTGTGACACAGAAATTTTATTTGTCATTTTGTGTTTCT 3299
Db 13031 TTTCTTATTTT--ATCGATTAAATGATTTCCGTTAAATCGACACAGCGCTACAGCAG 13088
|||||
Qy 3300 TTTGATATTTTGTGATCAATTAATTAATTTTCCGTTAAACAGACACAGCGCTACAGTAC 3359
Db 13089 TCACTTAAAGAGTACTAGCTTTGCTATGAGATATTTTG 13130
|||
Qy 3360 TCTTTTAAAGAGTTACAGTAGTTTTCGCTTCAAGATATTTTG 3401

RESULT 11

LOCUS AC006903 274626 bp DNA HTG 24-FEB-1999
DEFINITION Caenorhabditis elegans clone Y74C10, WORKING DRAFT SEQUENCE, 15
unordered pieces.

ACCESSION AC006903
NID 94263445
VERSION AC006903.1 GI:4263445
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 274626)

AUTHORS Waterston, R.H.
TITLE The sequence of Caenorhabditis elegans clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 274626)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2552: contig of 2552 bp in length
* 2553 2567: gap of unknown length
* 2568 5380: contig of 2813 bp in length
* 5381 5395: gap of unknown length
* 5396 8631: contig of 3236 bp in length
* 8632 8646: gap of unknown length
* 8647 12700: contig of 4054 bp in length
* 12701 12715: gap of unknown length
* 12716 18221: contig of 5506 bp in length
* 18222 18236: gap of unknown length
* 18237 23896: contig of 5660 bp in length
* 23897 23911: gap of unknown length
* 23912 31012: contig of 7101 bp in length
* 31013 31027: gap of unknown length
* 31028 38962: contig of 7935 bp in length

* 38963 38977: gap of unknown length
* 38978 46049: contig of 7072 bp in length
* 46050 46064: gap of unknown length
* 46065 61946: contig of 15882 bp in length
* 61947 61961: gap of unknown length
* 61962 84927: contig of 22966 bp in length
* 84928 84942: gap of unknown length
* 84943 99967: contig of 15025 bp in length
* 99968 99982: gap of unknown length
* 99983 136628: contig of 36646 bp in length
* 136629 136643: gap of unknown length
* 136644 176604: contig of 39961 bp in length
* 176605 176619: gap of unknown length
* 176620 274626: contig of 98007 bp in length.
Location/Qualifiers
1. 274626
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y74C10"

BASE COUNT 87518 a 49452 c 49335 g 88111 t 210 others
ORIGIN

Query Match 13.1%; Score 131; DB 19; Length 274626;
Best Local Similarity 82.1%; Pred. No. 5.67e-49;
Matches 238; Conservative 0; Mismatches 41; Indels 11; Gaps 7;

Db 61526 CAAATATCTCGTAGCGGAAACTACAGTAATTCCTTAAATGACTACTGTACACGGGTGT 61585
|||||
Cp 3401 CAAATATCTTGAAGCGGAAACTACTGTACTCTTAAAGAGTACTGTAGCGCTGTGT 3342
Db 61586 CGATTACGAGATCATCT--TTAATCGAT--AAAATATTAAGAAACACAAATGAC 61642
|||||
Cp 3341 CTGTTACGGAATATTTATTAATTTGATCAAAAATATCAAAAGAAACACAAATGAC 3282
Db 61643 AAAATAGTTTGTATTCGCTGCATTTTGAATATTAATTCGTTTGAATTCACGCCG 61702
|||||
Cp 3281 AAATTA--AT--TCGTTGTCACATTTGAGCATATATTCGTTTGAATTCACGCCG 3226
Db 61703 TAAATCGACCTACAGTAATATGATTAAGAAATTAATTTTTCGCTACGAGATAT 61762
|||||
Cp 3225 TAAATCGAC--ACTACAA--TAGTCATTTAAAGGGTACTGTATTTTCGCTACGAGATAT 3169
Db 61763 TTGCGGCTCAATATGTTGTGCAATACGCATTTTCGAGTTCAATATTC 61812
|||||
Cp 3168 TT-GCGGCTCAATATGTTGTGCAATACGCATTTTCGAGATTTTGTGTTG 3120

RESULT 12

LOCUS AC006751 340801 bp DNA HTG 23-FEB-1999
DEFINITION Caenorhabditis elegans clone Y39G10, WORKING DRAFT SEQUENCE, 29
unordered pieces.

ACCESSION AC006751
NID 94263187
VERSION AC006751.1 GI:4263187
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 340801)

AUTHORS Waterston, R.H.
TITLE The sequence of Caenorhabditis elegans clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 340801)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

Db 192245 GATTTCGAATGATTTATTTTCGAAAAATGACACGAAATTTATTTCTTACTTTT 192304
|||||
QY 3233 GATTTTCGAAACGAATATATCTCGAATGTGACACAGAA-TTTTAATTTGTCATTTT 3291
Db 192305 TGTGTTTCTTTTAAATATTTTATCCATTAAATGAATTCCTAAATCGACACGCGC 192364
|||||
QY 3292 TGTGTTTCTTTGATATTTTGAATCAATTAATAATATTTCCGTAACAGACACGCGC 3351
Db 192365 TACAGTAGTCATTCAGAAATTAAGTCTAGTTTTCACAGATATTTG 192414
|||||
QY 3352 TACAGTACTCTTTTAAAGAGTTACAGTAGTTTTCGCTTCAAGATATTTG 3401

RESULT 10
LOCUS CELM01H9 30522 bp DNA INV 30-SEP-1996
DEFINITION Caenorhabditis elegans cosmid M01H9.
ACCESSION U70853
NID 91572795
VERSION U70853.1 GI:1572795
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans strain=Bristol N2.
Caenorhabditis elegans
Eukaryotes; mitochondrial eukaryotes; Metazoa; Nematoda;
Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
Rhabditidae; Caenorhabditis.
REFERENCE
1 (bases 1 to 30522)
Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
Saunders,D., Showkeen,R., Smaildon,N., Smith,A., Sonhammer,E.,
Staden,R., Sulston,J., Thierri-Mieg,J., Thomas,K., Vaudin,M.,
Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
Wilkinson-Sproat,J. and Wohlman,P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans

TITLE
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE
AUTHORS 2 (bases 1 to 30522)
TITLE Pauley,A and Gattung,S.
JOURNAL The sequence of C. elegans cosmid M01H9
REFERENCE Unpublished (1996)
AUTHORS 3 (bases 1 to 30522)
TITLE Waterston,R.
JOURNAL Direct Submission
COMMENT Submitted (15-SEP-1996)
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: rwenematode.wustl.edu and jes@sanger.ac.uk

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is R08C7, 200 bp overlap; 3' cosmid is ZK180, 200 bp
overlap. Actual start of this cosmid is at base position 92 of
CELMO1H9; actual end is at 11962 of CELZK180

NOTES:

Coding sequences below are predicted from computer analysis, using
the program Genefinder(P. Green and L. Hillier, ms in preparation).
Location/Qualifiers
1..30522
/organism="Caenorhabditis elegans"

/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="M01H9"
complement(1636..3073)
/gene="M01H9.2"
complement(join(1636..1850,1909..2237,2298..2560,
2609..2685,2738..2872,2923..3073))
/gene="M01H9.2"
/note="weak similarity to mouse fibrillin-1 precursor
(GI:575510) and the Emericella nidulans anucleate primary
sterigmata coiled-coil protein (GB:X82289)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB09143.1"
/db_xref="PID:g1572797"
/db_xref="GI:1572797"

/translation="MLQLLFIFCSFTNASSCQDKSPYCNPNDCVPRPGYAMYCRK
TCGNACDFCEDSKFTTCSAERKDCDDMLSDYCPRLCGKCYAKLPDSKRTKTIPTQ
FKRNPAATSTTTTTRSPSRMKQPRMLPNGTFINPLPKYEKLDOTTYTIPHTVI
DYPIAHMEELMPEPORIWPEPEMRIEISVSPSYFVPOQVSOQPSWSPSLGL
SSLSDSRRAPONYFSSYQYPOYPDEKLTNTIAPNMQVLEPFLTPGQNDLNPSSMS
LNLGCKDKDEVICKHVTADCLSRPGYLYLKCPYCKNCSGYQCIDSIKIDCAEVK
AQGACLSVASEYCPRTCEYCNPPSDMAKTMDSCKDELETCEQLAESGACQHDR"
complement(9944..11767)
/gene="M01H9.3"
complement(join(9944..10045,10099..10309,10362..10510,
10745..10868,11739..11767))
/gene="M01H9.3"

/note="coded for by C. elegans cDNA yk151a7.5; coded for
by C. elegans cDNA yk184f9.5; coded for by C. elegans cDNA
yk91a11.5; coded for by C. elegans cDNA yk99b12.5; coded
for by C. elegans cDNA yk91a11.3; coded for by C. elegans
cDNA yk151a7.3"
/codon_start=1
/protein_id="AAB09144.1"
/db_xref="PID:g1572798"
/db_xref="GI:1572798"
/translation="MPLDPMDSLMSKAPQVSDLVQPNVTFKAVTGSBSPSASIP
IAGADDSIRPDYDTCLEMKPPTRSARVQKERTISMDSTSEIGALGKRFSTNEGASP
APTSPIDAOPTARMSISEMLFGSSPKSFGMESNTNTTTSNSGGDIGERKMSI
TDVSGWKQIDPRFKDMKHQSKIIGDDGISAAAGFKRSNYMKD"
complement(20695..23630)
/gene="M01H9.4"
complement(join(20695..20957,21298..21624,22484..22621,
22670..22751,23583..23630))
/gene="M01H9.4"

/codon_start=1
/evidence=not_experimental
/protein_id="AAB09145.1"
/db_xref="PID:g1572799"
/db_xref="GI:1572799"
/translation="MSETVSLKRPRSESDVTEEPDEKYSIHAGNHQOLMNPVAMGM
PNARGLQPNVGFPLMOYCOMIQMHQOMEYPMPIRKDNKDEANSASTPSSSR
QTLGAGAGPRIRVPQSIENPMQFGMPLSQQIIEEPSTPTMPROQSDPHTVH
MPSFORADEEITYVRIEGLSISRGREIKNTQPSIFGKHKTPTFFRFDSNMISTL
FSVIIPILVWELVFGVLAPDEVFMDVGEYNVPIDNRIVSPQSVRDVLVSSGDCG
RGELAVLS"
28085..29825
/gene="M01H9.1"
join(28085..28168,29021..29086,29201..29494,29745..29825)
/gene="M01H9.1"
/note="contains weak similarity to thioredoxin
redox-active center"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB09142.1"
/db_xref="PID:g1572796"
/db_xref="GI:1572796"

/translation="MAKNFSGTSLRLKDGTMVDAGEHLKRIVLYFSASWCGPCRO
FTPIMKELYOQIAATNPQIEVLISRDYMRFOLDIEYESHGCSGWVPLRDPITIEKCL
EKYDKALPSCRVVDFGNCIDANARQSVENVIFENWRNVDSWREFYREKYMALFL
NKWTEKGNLERIRA"

[illegible][illegible]

CP	3400	AAAAATATCTTGAAGCGAAACTACTGTAACTCTTTAAAGAGTACTGTGACCGCTGTGTC	3341
Db	14480	GATTTACGGAATCATTTATTTGATCGATAACACACAATAACAACAATCAAAAAACTA	14539
CP	3340	TGTTTACGGAATAATTTATTTAATTGATCAAA-A-ATATCA-AAAGAAACACAAAAATG	3284
Db	14540	ACGAAAAATTAAATTCGCTGCACACATTCGACACATAAATTAGTTTCATTAATCGAGCCCG	14599
CP	3283	AC--AAATTAAATTCGTTGTGCACATTCGACGATATATTTCGTTCCAAAAATCGAGCCCG	3226
Db	14600	TAAATGACACTAGCGTTACATTAGTCATTTAAGAATTACTGTGCTTTTCGCTACGACA	14659
CP	3225	TAAATCGACACTA-C---A-AT-AGTCATTTAAAGGCTTACTGTATTTTTTCGCTACGACA	3172
Db	14660	TACTTTGCGCGTCAAAATATGTTACGATATTACGCATTCCTCAGAATTAGTGTT	14711
CP	3171	TA-TTTGCGCGTCAAAATATGTTGAGTAATGCGCATTCCTCAGAATTAGTGTT	3121

RESULT	6
LOCUS	CEY32F6 187816 bp DNA HTG 09-NOV-1997
DEFINITION	Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from clone Y32F6, WORKING DRAFT SEQUENCE.
ACCESSION	AL008875
NID	G2612777
VERSION	AL008875.1 GI:2612777
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	Caenorhabditis elegans.
ORGANISM	Caenorhabditis elegans

REFERENCE AUTHORS TITLE	JOURNAL
1 (bases 1 to 187816) Sulston, J. Direct Submission	Nematode Sequencing Project Sanger Centre
Submitted (09-NOV-1997)	

COMMENT

IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments.

- * NOTE: This is a 'working draft' sequence.
- * This record will be updated with the finished sequence
- * as soon as it is available and the accession number will
- * be preserved.

FEATURES

FEATURES	Location/Qualifiers
source	1. .187816
	/organism="Caenorhabditis elegans"
	/db_xref="taxon:6239"
	/chromosome="V"
	/clone="Y32F6"
BASE COUNT	59977 a 33547 c 32230 g 59523 t 2539 others
ORIGIN	

Query Match	13.4%;	Score 134;	DB 19;	Length 187816;
Best Local Similarity	83.2%;	Pred. No. 1.51e-50;		
Matches	243;	Conservative	0;	Mismatches 37;
			Indels	12;
			Gaps	9;

[illegible]

CP 3283 AC- AAATTAAATTCGTTGTGCACAATTGCAGCATATATTGTTTGGAAATCGAGCCCG 3226

Db 14610 TAAATGGACACTAGCGTTACATTAGTCATTTAAAGAATTACTGTCGTTTTCGCTACGACA 1466

CP 3225 TAAATCGACACTA-C---A-AT-AGTCATTTAAAGGGTTACTGTATTTTTCGCTACGAGA 3172

Db 14670 TACTTGGCGCGTCAATATGTTTCAGTATTACGCATTCAGAAATTAGTGT 14721

CP 3171 TA-TTGGCGCGTCAATATGTTGAGTAATGCGCATTCACAGATTTAGTGT 3121

RESULT	7
LOCUS	AC006670 36454 bp DNA HTG 23-FEB-1999
DEFINITION	Caenorhabditis elegans clone K07A3, WORKING DRAFT SEQUENCE, 1
ACCESSION	unordered pieces. AC006670
NID	94263268
VERSION	AC006670.1 GI:4263268
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	Caenorhabditis elegans.

REFERENCE	1 (bases 1 to 36454)
AUTHORS	Waterston, R.H.
TITLE	The sequence of <i>Caenorhabditis elegans</i> clone
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 36454)
AUTHORS	Waterston, R.H.
TITLE	Direct Submission
JOURNAL	Submitted (23-FEB-1993) Genome Sequencing Center Washington

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES

```

/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="K07A3"
BASE COUNT      11652 a      6685 c      7078 g      11039 t
ORIGIN

```

Query Match	13.2%;	Score 132;	DB 19;	Length 36454;
Best Local Similarity	79.7%;	Pred. No. 1.70e-49;		
Matches 231; Conservative	0;	Mismatches 51;	Indels 8;	Gaps 6.

Db	1911	CAAAATATCTCGTAGTGAAAACTACAGTAATCTTTGATGACTACTGTAGCGCTGTGT	1970
Cp	3401	CAAAATATCTTGAAGCGAAAACTACTGTAACTCTTTAAAGAGCTACTGTAGCGCTGTGT	3342
Db	1971	CGATTTACGGAAATCATTTATTTAATGGATAAAATATTTAAAGAAAAACAAAAACTAAAG	2030
Cp	3341	CTGTTTACGGAATTAATTTATTTAATTTGATCAAAAAATATCAAAAGAAACACAAAAATGAC	3282
Db	2031	AAAAATAAATTTGTTGTCATTTTTCGAAATAATAATTCATTTTCGAAATCTCGCCCGTAA	2090
Cp	3281	AAATTTAAATTT-CGTTGTACACAATTCGAGCATATATTCGTTTCGAAAAATCGAGCCCGTAA	3223
Db	2091	ATCGACACAAGAGCTACAGTAATCATTTTAAGAATTACTGTAGATTTTCGCTACGAGATAT	2150
Cp	3222	ATCGACACTA--C-A-A-TAGTCATTTAAAGSGTTACTGTATTTTTCGCTACGAGATAT	3169
Db	2151	TTTGCGGCTCAATATGTGTTGCGCAGTACGTATCTCAGAAATTTTTCGCTC	2200
Cp	3168	TT-GCGGCTCAATATATGTTGAGTAATGCGCAATCTTCAGAAATTTAGTGTTC	3120


```

/db_xref="PID:g3880758"
/db_xref="GI:3880758"
/translation="MSATEKDKTPLLGVSQPISTAHDPDSKDIAQTLLIETLKKESGY
EPKEETEQRMEVLRLNRLVKEKQNTAMKIPNGEVNAGKLTFTGSGYGVHSSG
ADIDITLAVVPRHIDRSDFTSTKEMLNNDPNTLHGVEAFVPMKLYSGVELDIL
FARLAKVEYDQELSDNLRSLNGSVSLNGCRVAEQLKLVPRQKECVTLRLAI
KLWAKNNGIYSNMGFFGGITWAILVARACQLYPNAPSRLVHRMFIESTWPHPV
VLNEMNDRNDIPTLCELWDPDRNTDREHMPITPAFPEONSTHNTSTATVIVK
NEICEALEICRDISGSKWTALFEVNEFSRYKHFIALIMAAPNEEELNNGFLES
RIRLVOSLERMODIIAHNDPNKPKSPNAKEDVNPENKRVTVWFIQLEFAEHAATL
DLTNEIQRKTVELQASNVKIGENCQVQIDMEFYKRNLSLQVISAADLRGRKWK
VPIATNTSVSSSTPRSVRTSTSVPTPTGLAAPKTPUSASVATNEPDSSTNGT
PLSRKSMDEESTVTTSQISDESVPKKTTRDDTLEENRYSMVVEVSNVVEQRTKVY
QEIVDLQADNGLNTSNGLEASQKMEYQSV"
join(complement(271180.1:2698..2908),
complement(271180.1:1785..2041),
complement(271180.1:1541..1741),
complement(271180.1:1152..1497),
complement(271180.1:115..1106),complement(27038..27236),
complement(26919..26986),complement(26026..26313),
complement(25704..25934),complement(25543..25653),
complement(24197..24835),complement(19242..19310))
/gene="F22E12.1"
/note="similar to BPTI/KUNITZ inhibitor domain; CDNA EST
EMBL:D68293 comes from this gene; CDNA EST yk448h4.5 comes
from this gene; CDNA EST yk249e6.5 comes from this gene;
CDNA EST yk448h4.3 comes from this gene"
/codon_start=1
/protein_id="CA16311.1"
/db_xref="PID:e1350302"
/db_xref="PID:g3880760"
/db_xref="GI:3880760"
/db_xref="SPTREMBL:O45916"
/translation="MLVLILLFGCTAVNSQQLSNPKCNHWPDRGTCELDFHVKKWY
DRYDHCRRFFYGGCEGNGENRFDLSLECCSCHYQVPTNRDRCFOPHDGNCYADIER
WFEDEKNCOCVSMWSGCGGNSNIYYSYNHCMLICGEYAEHGPIDEXYGRQMRSM
SAESLHFNPNTESTHOYSEEPYVQVPLENSYIDSHPRRYGNSDMNDKMLTINIS
RDDGSPFHAAPAVYVDMFSRAAQSFQADGLTHRYDSEPRPMOVLQDQVNWQIQE
QASAOQNGFMKRRFKMOKKLLPRMHLGTPRASQSNRYRIQLDSEINENRPTVOY
VREQDAHVOQHNOISQSLNEAEETORRIODTMRNRYFEQARYQRPVQMPPPPPO
QNSPSVNEHYHPELVNQFSLDEHEKALRKLEAOPDHIITLPHLETVRNPDGSN
VVRQRIKWTPYKGARIPDEHVPVPPVAQLPRLPMQVIOPILTQAVYTYTEYIPL
ITVPTHEISRLKHEEQKRYKEDLEKROEQKQOEKKKIQELKAAQOEREREM
VEREKAENDRLKYENEMKQRAEYQARMRLTPTPTTVEETTEYHVPYPTLPVN
EYVYDAHIRKPIPTSKSIAEPHFHPIIENVTPASLSLEDIAMDORSDEYEDYV
PMDPPEAPPLTTRPPVTRNTRPPRPQIPQRPSPINGMPSIVOBSSGKSKL
FAPPSSEYTOEVDFFPWNLLFLLSLVSLAAGASNSVNTGWGIQYDEDWGYDIR
TNAHTFWLWYAAKPAANSQRLFLMLQGGPGSSSGFNGFEETGPKTLNGSDNPATLQ
VADMYVVDNPVAGFSYVDKSAYTEITQIGKDLAWLRKFLALHPEYRTPYIFC
ESYGGKMSAQFAKITDSIKAGSLQNFRAVALGDSWISAMDYVNTWGPYLYANSFLD
DHQNTVNAEARCOALVDOCKATNCWGNMNLISVETNDWSYNIILKGDITDW
SSSAMSRSNRVMTSTRRLYNRYAPQNLDSLNYMDTVVRKKLGIIIPDKVRGGQAGDV
FSYQGDFTMTPIWSTVDLKDGNVIVYNGNEDLICNTMGTAAMVNRILTMDGAATEN
STRHSEKTSQSFPLAGYKYKYLQFWILRAGHAYADTPESAIFMLKAVYKQYNA"
join(complement(6241..6293),complement(6089..6195),
complement(5656..5745),complement(5110..5281),
complement(4965..5062),complement(4090..4182),
complement(2542..2748),complement(2003..2488),
complement(1841..1944),complement(1734..1787),
complement(723..788),complement(325..670),
complement(105..141),complement(273972.1:25167..25503),
complement(273972.1:25055..25113),
complement(273972.1:24908..25008),
complement(273972.1:24744..24861),
complement(273972.1:24468..24612))
/gene="F15H10.4"
join(complement(6241..6293),complement(6089..6195),
complement(5656..5745),complement(5110..5281),
complement(4965..5062),complement(4090..4182),
complement(2542..2748),complement(2003..2488),
complement(1841..1944),complement(1734..1787),
complement(723..788),complement(325..670),
complement(105..141),complement(273972.1:25167..25503),

```

Note: remainder of annotations omitted.

```

Query Match          13.4%; Score 134; DB 21; Length 27236;
Best Local Similarity 83.2%; Pred. No. 1,51e-50;
Matches 243; Conservative 0; Mismatches 37; Indels 12; Gaps 9;

Db 8976 AAAATATCTCGTAGCGAAACTACAGTAACTATTAAATGACTACTGTAGCGGTGTC 9035
||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1
Cp 3400 AAAATATCTGAAGCGAAACTACTGTAACCTTTAAAGAGTACTAGCGGTGTC 3341
||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1
Db 9036 GATTTCGGAATCATTTATGATCGATACACACAATAACAAACAATAACAAACTA 9095
||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1
Cp 3340 TGTTCACGGAATAATTAATTAATGATCAAA-A-ATATCA-AAAGAAACACAAAAATG 3284
||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1
Db 9096 ACGAAATTAATTTCCGTCGCCACAAATTCGACACATAAATTAGTTCAATATCGAGCCCG 9155
||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1
Cp 3283 AC-AAATTAATTCGTTGTCACAAATTCGAGCATATATTCGTTGCAAAATCGAGCCCG 3226
||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1
Db 9156 TAAATGACACTAGCGTTACATTAATGATTAAGAAATTAATGATGCTGTTTCGCTAGACA 9215
||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1
Cp 3225 TAAATCGACACTA-C--A-AT-AGTCATTTAAAGGTTACTGTATTTTCGCTAGAGA 3172
||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1
Db 9216 TACTTGGCGCTCAATATAGTTCAGTATTAGCGCATCTCAGAAATTAGTGT 9267
||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1
Cp 3171 TA-TTGGCGCGTCAATATAGTGAATGAGTATGCGCATCTCAGAAATTAGTGT 3121
||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1

```

```

RESULT 5
LOCUS CEK07H11 36879 bp DNA HTG 14-OCT-1998
DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
clone K07H11, WORKING DRAFT SEQUENCE.
ACCESSION Z81566
NID 93377978
VERSION 281566.1 GI:3377978
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 36879)
AUTHORS Steward,C.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jesesanger.ac.uk or rwenematode.wustl.edu
COMMENT On Aug 3, 1998 this sequence version replaced gi:1695068.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

FEATURES
    source
        1..36879
            /organism="Caenorhabditis elegans"
            /db_xref="taxon:6239"
            /chromosome="V"
            /clone="K07H11"
BASE COUNT 11032 a 6897 c 6237 g 11112 t 1601 others
ORIGIN

```

```

Query Match          13.4%; Score 134; DB 19; Length 36879;
Best Local Similarity 83.2%; Pred. No. 1,51e-50;
Matches 243; Conservative 0; Mismatches 37; Indels 12; Gaps 9;

Db 14420 AAAATACTCGTAGCGAAACTACAGTAACTATTAAATGACTACTAGCGGTGTC 14479

```


Db 240091 ATGCGGGCTCCGACGATGCTTGGCCGGTGACATTTGACACTCGAATTGACAGCATTCG 240150
|||||
Cp 2921 ATGCGGGCTCCGACGATGCTTGGCCGGTGACATTTGACACTCGAATTGACAGCATTCG 2862
Db 240151 AGTCAACACTGGAAAAAGGATATTTGAAGGAGATGTGTGTGTGTGTGACAACTA 240210
|||||
Cp 2861 AGTCAACACTGGAAAAA-GGGATATTGGAAGGAGATGTGTGTGTGTGTGACAACTA 2803
Db 240211 TTTTCGGCAATCTACAAATGCGCTTTTGGCCGACGAAATAGTGTAAATAAATAAT 240270
|||||
Cp 2802 TTTTCGGCAATCTACAAATGCGCTTTTGGCCGACGAAATAGTGTAAATAAATAAT 2743
Db 240271 AATCACTCGAGCTTTCAGTAATAATAATATTTCACTTCAA 240312
|||||
Cp 2742 AATCACTCGAGCTTTCAGTAATAATAATATTTCACTTCAA 2701

RESULT 4
LOCUS CEY32F6A 27236 bp DNA INV 23-NOV-1998
DEFINITION Caenorhabditis elegans cosmid Y32F6A, complete sequence.
ACCESSION AL021474
NID 92815144
VERSION AL021474.1 GI:2815144
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 27236)
Barlow, K.
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (23-JAN-1998) Louis, MO 63110, USA. E-mail:
REFERENCE jesus@sanger.ac.uk or rwenematode.wustl.edu
2 (bases 1 to 27236)
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A.,
Saunders, D., Showkneen, R., Smaildon, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Wohlman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
COMMENT Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:-
http://webace.sanger.ac.uk/cgi-
bin/display?db=wormace&class=Sequence &object=Y32F6A
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone Y32F6A.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone F22E12 is at 12868 in this sequence. The
true right end of clone F15H10 is at 104 in this sequence. The
start of this sequence (1. 104) overlaps with the end of sequence

273972.
The end of this sequence (27123. 27236) overlaps with the start of
sequence 271180.
FEATURES
source
location/Qualifiers
1. 27236
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="V"
/clone="Y32F6A"
9724. 13659
/gene="Y32F6A.1"
join(9724. 9844,9893. 9984,10030. 10119,10163. 10262,
10312. 10479,10526. 10628,10671. 10884,10933. 11025,
13609. 13659)
/gene="Y32F6A.1"
/note="cDNA EST YK314d7.3 comes from this gene; cDNA EST
YK314d7.5 comes from this gene"

gene
CDS

gene
CDS

gene
CDS

/complement(11074. 13503)
/gene="Y32F6A.2"
complement(join(11074. 11169,11210. 11529,11820. 11958,
12005. 12236,12284. 12585,12632. 13321,13369. 13503))
/gene="Y32F6A.2"
/note="similar to Sodium:neurotransmitter symporter family
(2 domains)"
/codon_start=1
/protein_id="CAA16307.1"
/db_xref="PID:e1350298"
/db_xref="PID:g3880756"
/db_xref="GI:3880756"
/db_xref="SPTREMBL:O45915"
/translation="MASSSDFKOLKOTRKSEETTRASASNETVNDNRSTTEIPI
KTSNATAISNIGFESVTNLALENIGLDTKEGPTWTSKWEAITATLSFTCSGNIF
FPYLCGYGWPFPYQETFCYFIAVPLYLETALGOYASASPLSTFRMAPAMAGLSA
GMCIFMFRYLSBPNESEFQNPQFRTISLSOWAIDLTFTTHASQSTWSTWESCOE
SOSGDYCVNYSKLSKCTWQIQNSDEKEDYQEMLAITRGFOAKSPFMSFVHGLMYR
SITMDWEPSPVTSIICAVCLWIIIGISIGSKVLGRGTGIVALTLLGLTLLSYG
MSLNDTNTVFAFYQSEGEDKMMWMSWADAHAHALRALNVGGGIGQKFASSLNNEH
NKIHRDVLISIVSEVFYICTGLSFMFAEIGRLYPDLSESRIOLYATPMIESV
ISEILTNSMSGSIWFLWTLAACSIOGSIYSIWEMKPKRQYLSLESLPIHNNL
KOKFSPKNYFRNIRAMIGGHGHPNFWMLNVLVISPILLVTFGCIVATFGQKAFN
ENSIVSDPIGSLVMPCIFVILYFLRDEYDRKMEPFVVMIRATGDMGPMNPEDRR
NAKFERQLRVRY"
14825. 17771
/gene="Y32F6A.3"
join(14825. 15049,15355. 16243,16293. 16618,17244. 17771)
/gene="Y32F6A.3"
/note="similar to POLY(A) POLYMERASE (EC 2.7.7.19) (PAP)
(POLYNUCLEOTIDE ADENYLYLTRANSFERASE); cDNA EST YK317f12.3
comes from this gene; cDNA EST YK317f12.5 comes from this
gene; cDNA EST EMBL:D34422 comes from this gene; cDNA EST
EMBL:D37481 comes from this gene; cDNA EST EMBL:D37455
comes from this gene; cDNA EST EMBL:D67432 comes from this
gene; cDNA EST EMBL:D67884 comes from this gene; cDNA EST
YK239f3.5 comes from this gene; cDNA EST YK250b11.5 comes
from this gene; cDNA EST YK276e1.5 comes from this gene;
cDNA EST YK308b2.5 comes from this gene; cDNA EST
YK349g6.5 comes from this gene; cDNA EST YK356g6.5 comes
from this gene"
/codon_start=1
/protein_id="CAA16309.1"
/db_xref="PID:e1350300"


```
Cp 3460 GAAAAAGCGTTTGGCACATGTTAGATGATTTTTCAAAATGTTAAATTCCTTTC 3401
Db 9954 AAAATATCTGGAAGCGAAACACTGTAACCTTTAAAGAGTACTGTAGCGGTGTC 10013
Cp 3400 AAAATATCTGGAAGCGAAACACTGTAACCTTTAAAGAGTACTGTAGCGGTGTC 3341
Db 10014 TGTTCAGGAAATATTTATTTATGATCAAAAATATCAAAAGAAACACAAAAATGACA 10073
Cp 3340 TGTTCAGGAAATATTTATTTATGATCAAAAATATCAAAAGAAACACAAAAATGACA 3281
Db 10074 AATTAATTCGTTGTCAACAATTCGACATATATTCGTTGAAAATCGAGCCCGTAAT 10133
Cp 3280 AATTAATTCGTTGTCAACAATTCGACATATATTCGTTGAAAATCGAGCCCGTAAT 3221
Db 10134 CGACACTACATAGTCAATTTAAAGGTTACTGTATTTTCGCTACGAGATATTTGCGCG 10193
Cp 3220 CGACACTACATAGTCAATTTAAAGGTTACTGTATTTTCGCTACGAGATATTT-GCGCG 3162
Db 10194 TCAATATGTTGAGTAATGCGCATCTCGAATTTAGTTCGCATCAACATACAAGTT 10253
Cp 3161 TCAATATGTTGAGTAATGCGCATCTCGAATTTAGTTCGCATCAACATACAAGTT 3102
Db 10254 GCGTGGGAAATGCGTTGACTGGAGAGATGAATATTTGTCGATCCGATGAATGAGTG 10313
Cp 3101 GCGTGGGAAATGCGTTGACTGGAGAGATGAATATTTGTCGATCCGATGAATGAGTG 3042
Db 10314 CACGCGATGAGAAAGAGATCTTCTGCTGAGTAGATATCTGATTAAGAGTGAATGATG 10373
Cp 3041 CACGCGATGAGAAAGAGATCTTCTGCTGAGTAGATATCTGATTAAGAGTGAATGATG 2982
Db 10374 ACACGTGAAGAGACGCTGTACAGGTGAAGTCCGGTGGTGAAGTGAAGTGAAGTGA 10433
Cp 2981 ACACGTGAAGAGACGCTGTACAGGTGAAGTCCGGTGGTGAAGTGAAGTGAAGTGA 2922
Db 10434 ATGCGCGGCTCCGACGATGCGTTCGCGGTGACATTTGACACATTCGACACGATTCG 10493
Cp 2921 ATGCGCGGCTCCGACGATGCGTTCGCGGTGACATTTGACACATTCGACACGATTCG 2862
Db 10494 AGTCAACACTGGAAGAAAGGATATTTGAAGGAGATGTGTGTGTGTGTGACACAGTA 10553
Cp 2861 AGTCAACACTGGAAGAAAGGATATTTGAAGGAGATGTGTGTGTGTGTGACACAGTA 2803
Db 10554 TTTTCGGCAATCTACAAATGGCTTTTGGCCGACGAAATTAAGTGTAAAAATTAAT 10613
Cp 2802 TTTTCGGCAATCTACAAATGGCTTTTGGCCGACGAAATTAAGTGTAAAAATTAAT 2743
Db 10614 AATCACTCGAGCTTTCAGTAATAATAATATTTCACTTCAA 10655
Cp 2742 AATCACTCGAGCTTTCAGTAATAATAATATTTCACTTCAA 2701

RESULT 3
LOCUS CEY67H2 314495 bp DNA HTG 04-MAR-1999
DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
clone Y67H2, WORKING DRAFT SEQUENCE.
ACCESSION AL022475
NID 94469034
VERSION AL022475.3 GI:4469034
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 314495)
AUTHORS McMurtry,A.
TITLE Direct Submission
JOURNAL Submitted (04-MAR-1999) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or twenematode.wustl.edu
COMMENT On Mar 22, 1999 this sequence version replaced gi:4468145.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
```

```
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
    source
        1..314495
        /organism="Caenorhabditis elegans"
        /db_xref="taxon:6239"
        /chromosome="IV"
        /clone="Y67H2"
BASE COUNT 96542 a 54551 c 52760 g 97027 t 13615 others
ORIGIN

Query Match 93.0%; Score 930; DB 19; Length 314495;
Best Local Similarity 98.9%; Pred.No. 0.0e+00;
Matches 991; Conservative 0; Mismatches 1; Indels 10; Gaps 9;

Db 239319 AAATGATGCGCTTTTAC-ATTTTCTGCACA-CATGAATAGGGGAAATATATATAA 239376
Cp 3700 AAATGATGCGCTTTTACTATATTTTCTGCACAAACATGATAGGGGAAATATATATAA 3641
Db 239377 AAT-GC-ATTTTCTGCAT-CATATGAACAA-CCACATGATTTTTCGTTGAGCAATTC 239432
Cp 3640 AATGCTATTTTCTGCATCATATGAACAAACACATGATTTTTCGTTGAGCAATTC 3581
Db 239433 AAAAGTAGAGCATCTAGAACGAA-CCAAATTTCTACAGATATTTACATTAATGA 239490
Cp 3580 AAAAGTAGAGCATCTAGAACGAAACCCAAATTTCTACAGAAATTCACATTAATGA 3521
Db 239491 TAATATATGTTAATAAGCATATCTTGAATGAAAAAATCAAAAAATCTGCGAACTT 239550
Cp 3520 TAATATATGTTAATAAGCATATCTTGAATGAAAAAATCAAAAAATCTGCGAACTT 3461
Db 239551 GAAAAAGCGTTTGGCACATGTTAGATGTTTTTTTCAAAATGTTAAATTCCTTTC 239610
Cp 3460 GAAAAAGCGTTTGGCACATGTTAGATGTTTTTTTCAAAATGTTAAATTCCTTTC 3401
Db 239611 AAAATATCTGGAAGCGAAACACTACTGTAACTCTTTAAAGAGTACTGTAGCGTGTGTC 239670
Cp 3400 AAAATATCTGGAAGCGAAACACTACTGTAACTCTTTAAAGAGTACTGTAGCGTGTGTC 3341
Db 239671 TGTTCAGGAAATTAATTTATTAATGATCAAAAATATCAAAAGAAACACAAAAATGACA 239730
Cp 3340 TGTTCAGGAAATTAATTTATTAATGATCAAAAATATCAAAAGAAACACAAAAATGACA 3281
Db 239731 AATTAATTCGTTGTCACAATTCGACATATATTCGTTGAAAATCGAGCCCGTAAT 239790
Cp 3280 AATTAATTCGTTGTCACAATTCGACATATATTCGTTGAAAATCGAGCCCGTAAT 3221
Db 239791 CGACACTACAATAGTCAATTTAAAGGTTACTGTATTTTCGCTACGAGATATTTGCGCG 239850
Cp 3220 CGACACTACAATAGTCAATTTAAAGGTTACTGTATTTTCGCTACGAGATATTT-GCGCG 3162
Db 239851 TCAATATGTTGAGTAATGCGCATCTCAGAAATTTAGTGTGCAATCAACATACAAGTT 239910
Cp 3161 TCAATATGTTGAGTAATGCGCATCTCAGAAATTTAGTGTGCAATCAACATACAAGTT 3102
Db 239911 GCGTGGGAAATCGTTGACTGGAAGAGATGAATAATTTGTGTCATCCGATGAATGAAGTG 239970
Cp 3101 GCGTGGGAAATCGTTGACTGGAAGAGATGAATAATTTGTGTCATCCGATGAATGAAGTG 3042
Db 239971 CACGCGATCGAGACGAGATCTTGTCTTGAAGTATATCTGATTAAGAGTGAATGATG 240030
Cp 3041 CACGCGATCGAGACGAGATCTTGTCTTGAAGTATATCTGATTAAGAGTGAATGATG 2982
Db 240031 ACACGTGAAGAGACGCTGTACGCGTGAAGTCCGGTGGTGAAGTGAAGTGAAGTGAAGTGA 240090
Cp 2981 ACACGTGAAGAGACGCTGTACGCGTGAAGTCCGGTGGTGAAGTGAAGTGAAGTGAAGTGA 2922
```


REFERENCE
AUTHORS

2 (bases 1 to 39908)
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,
Saunders, D., Shownkeen, R., Smaildon, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Wohlman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans

TITLE

JOURNAL
MEDLINE
COMMENT

COMMENT

Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:-
http://webace.sanger.ac.uk/cgi-
bin/displaydb=wormacefclass=Sequence &object=C48D1
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
This sequence is the entire insert of clone C48D1. The true right
end of clone F58D2 is at 18510 in this sequence. The start of this
sequence (1..100) overlaps with the end of sequence Z81093.
The end of this sequence (33020..39908) overlaps with the start of
sequence Z82274.

Location/Qualifiers
1..39908
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="C48D1"
complement(994..4439)
/gene="C48D1.1"
complement(join(994..1020,1345..1951,2283..2580,
2629..2765,2882..3147,3615..3739,4089..4331,4382..4439))
/gene="C48D1.1"
/note="similar to C.elegans ZYG-11 like protein; cDNA EST
EMBL:Z14909 comes from this gene; cDNA EST EMBL:T01768
comes from this gene"
/codon_start=1
/protein_id="CAB02846.1"
/db_xref="PID:e1344741"
/db_xref="PID:g3875048"
/db_xref="GI:3875048"
/db_xref="SPTREMBL:O02227"

FEATURES
source

gene
CDS

gene
CDS
complement(join(6258..6349,6876..7065,7420..7547,
7602..7815,8728..9059,10248..10501,10780..10925,
10989..11120))

gene
CDS

/gene="C48D1.2"
/note="similar to Caspase recruitment domain, ICE-like
protease (caspase) p20 domain., ICE-like protease
(caspase) p10 domain.; cDNA EST EMBL:D75533 comes from
this gene; cDNA EST yk198f10.5 comes from this gene; cDNA
EST yk491a5.5 comes from this gene"
/codon_start=1
/protein_id="CAB02848.1"
/db_xref="PID:e1344743"
/db_xref="PID:g3875050"
/db_xref="GI:3875050"
/db_xref="SPTREMBL:O02229"
/translation="MRDRRLRLRNIMFSSHLKVDLEVLIAKQVLSNDGDMIN
SCGYAREKREIVKAVQQRGVDAFYDALRSTGHEGLAEVLEPLARSDSNAVEFE
CPMSPAHRRSRALSPAGYTSPTRVHRDSVSSVSFTSYQDYSRARSRSRALHSS
DRHNSPPVNAFPSPQSSANSFTGSSLSGSSSRMSFSKASGPTQYIFHEDMNF
VDAPITSRVFEDEKTYRNFSRGMCLINNEHEFQMPTRNGTKADKDLTNLFRMG
YTVICRDNLTGRGMLTIRDFAKHSHGDSAILVILSHGEENVIIIGVDIPISHEIY
DLNANANAPRLANKPKIVFVQACRGERDNGEPVLDSVDGVPALRGWDNRDGLFN
FLGCVBPQVQVWRKKPSQADLILAYATTAQYVSWRNSRGSWFIAVCVEFTHAKD
MDVVELLLEVNKKVAYDIPPAQKVLVLAGSTKLCRLKFTRLSLPN"
complement(16940..21616)
/gene="C48D1.3"
complement(join(16940..17126,18419..18597,18985..19442,
19553..19879,19925..19961,20333..20481,20789..20987,
21037..21152,21373..21616))
/gene="C48D1.3"
/note="Similarity to Salmonella sodium/proline symporter
(SW:PUTP_SALTY); cDNA EST EMBL:C13840 comes from this
gene; cDNA EST EMBL:C11562 comes from this gene"
/codon_start=1
/protein_id="CAB02847.1"
/db_xref="PID:e1344742"
/db_xref="PID:g3875049"
/db_xref="GI:3875049"
/db_xref="SPTREMBL:O02228"
/translation="MADLGLIAIVFFVYLIVVGITWAGRKSSEKLESEGAATEE
VMLAGRNIGTLVGIETMTATWVGAYINGTAELNGGLGQAPVGAISLYMGLL
FAKMRREEGYITMLDPFOFWNLELIFGRTEFNPRKIFLQTIIEILDFQHKYG
ORIGLMYVPALLGETFWTAAILALGATLSVILGDMNASVTLASCIAYFTTGGY
YAAVYTDVQLFCIFVGLLILGLVQNRPNRFEKTESLMIDMLLVFGIGIPQVYFQ
VLSKTAHGAQTLSEFVAGVCIILMAIPALIGAIARNTDWMTDLSFPMNNGRVESIP
PDKRNWVPLVFQYLTPRWAFIGLAVSAVMSADSVLSAASFANHTMKLIRP
HASEKEVILVMRIAILICVGMATIMALTIOSYGLMYLCADLVYILFPOLCVYMP
RSNTYGLAGYAVGLVRLIGGEPLVSLPAFFHYPMYTDGVOYEPFRTTAMLSMATI
YIVSIQSEKLFKSGRLSPEDWMCVNVNIPIDHVPPLSDVSFAVSEETLNMKVECDGM
QEPOLQTEHRLQYIRTNRLMKIHYYIHRTKVLIPQIAIKRDELVPRVHLISSQIFF
CIFESFR"

BASE COUNT 13182 a 6945 c 7146 g 12635 t
ORIGIN

Query Match 93.0%; Score 930; DB 21; Length 39908;
Best Local Similarity 98.9%; Pred. No. 0.00e+00;
Matches 991; Conservative 0; Mismatches 1; Indels 10; Gaps 9;

Db 9662 AATGCATGCGCTTTTAC-ATTTTCTGCACA-CATGAATAGGGAATATATAAA 9719
|||||
Cp 3700 AATGCATGCGCTTTTACATTTTCTGCACAACATGATAGGGAATATATAAA 3641
|||||

Db 9720 AAT-GC-ATTTTCTCAT-CATATGACAA-CCACATGATTTTCTGAGCAATTC 9775
||| |||
Cp 3640 AATGCTATTTTCTCATATGACAAACACATGATTTTCTGAGCAATTC 3581
|||||

Db 9776 AAAAGTAGACATAGAGGAAAA--CCAAAATTTCTACAGATATTCACATTAATGA 9833
|||||
Cp 3580 AAAAGTAGACATAGAGGAAAAATTTCTACAGAAATTCACATTAATGA 3521
|||||

Db 9834 TAATATATGTGTAATAGCATATCTGAATGAAAAAATCTGGAACCTT 9893
|||||
Cp 3520 TAATATATGTGTAATAGCATATCTGAATGAAAAAATCTGGAACCTT 3461
|||||

Db 9894 GAAAAAGCGTTTGGACATGTTGATTTTCTCAAAATGTTAAATCTTTTC 9953
|||||

/protein_id="AAA27982.1"
/db_xref="PID:g456417"
/db_xref="GI:456417"
/translation="MMRDRSLERINIMFSSHLKVDLEVLIAKOVLSNDGDMINSCGIVREKREIVKAVQRGDVAFDAFYDALRSTGHEGLAEVLEPLARSDNAVEFECMPSPASHRRSRALSPAGYTSPTVRHSDSVSSSTSYQDIYSRARSRSRALHS SDRHNSSPVNAFPSPSSANSFTGCCSLGYSSSRNSFSKASGPTYIFHEDMNFVDAPITSRVDEKMLTYRNFSSPRMCLINNEHFEDMPTRNGTKADKDLTNLFCMGYTVICKDNLTRGMLLTIRDFAKHSHGDSAILVILSHGEENVIIIGVDDIPISTHEIYDLNMANAPRLANKPKIVFVQACRGERDNGFPVLDSYDVPALRGWDRDPLFNFGLCVROVOQVWRKKPSQADILIRYATTAQYVSWRNSARGSWFIQAVCEVSTHAKDMDVVELTEVNKKVACGFQTSQGSNILKQMPENTSRLLKFFYFWEARNASV"

intron 2367. .2429
/gene="ced-3"
/number=1
exon 2430. .2575
/gene="ced-3"
/number=2
intron 2576. .2853
/gene="ced-3"
/number=2
exon 2854. .3107
/gene="ced-3"
/number=3
intron 3108. .4302
/gene="ced-3"
/number=3
exon 4303. .4634
/gene="ced-3"
/number=4
intron 4635. .5546
/gene="ced-3"
/number=4
exon 5547. .5760
/gene="ced-3"
/number=5
intron 5761. .5814
/gene="ced-3"
/number=5
exon 5815. .5942
/gene="ced-3"
/number=6
intron 5943. .6297
/gene="ced-3"
/number=6
exon 6298. .6537
/gene="ced-3"
/number=7
intron 6538. .7012
/gene="ced-3"
/number=7
exon 7013. .7652
/gene="ced-3"
/number=8
BASE COUNT 2429 a 1455 c 1271 g 2498 t
ORIGIN

Query Match 100.0%; Score 1000; DB 21; Length 7653;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2701 TTGAAGTGAATATATTTATTACTGAAAGCTCGAGTGATTATTTTAACTA 2760
|||||
Qy 2701 TTGAAGTGAATATATTTATTACTGAAAGCTCGAGTGATTATTTTAACTA 2760
Db 2761 ATTTTCGTGGCGCAAAAGCCATTTTGTAGATTGCGGAAATCTTGTACACACACAC 2820
|||||
Qy 2761 ATTTTCGTGGCGCAAAAGCCATTTTGTAGATTGCGGAAATCTTGTACACACACAC 2820
Db 2821 ACACACATCTCTCAATATATCCCTTTTCCAGTGTGACTCGAATGCTGTGAATTGA 2880
|||||
Qy 2821 ACACACATCTCTCAATATATCCCTTTTCCAGTGTGACTCGAATGCTGTGAATTGA 2880

Db 2881 GTGTCCAATGTCACCGGCAAGCCATCGTCGGAGCGCGGCATTGAGCCCCCGCGCTACAC 2940
|||||
Qy 2881 GTGTCCAATGTCACCGGCAAGCCATCGTCGGAGCGCGGCATTGAGCCCCCGCGCTACAC 2940
Db 2941 TTCACCGAACCAGGTTCCACCGTGACAGCGTCTCTTCAAGTGTCAATCATCTTCTATCA 3000
Qy 2941 TTCACCGAACCAGGTTCCACCGTGACAGCGTCTCTTCAAGTGTCAATCATCTTCTATCA 3000
Db 3001 GGATATCTACTCAAGAGCAAGATCTCGTTCGTGATCGCGTGCACCTTCATTCATCGGATCG 3060
Qy 3001 GGATATCTACTCAAGAGCAAGATCTCGTTCGTGATCGCGTGCACCTTCATTCATCGGATCG 3060
Db 3061 ACACAAATTATTCATCTCTCCAGTCAACGCAATTTCCAGCCCAACCTTGTAATGTGATGCG 3120
Qy 3061 ACACAAATTATTCATCTCTCCAGTCAACGCAATTTCCAGCCCAACCTTGTAATGTGATGCG 3120
Db 3121 AACACTAAATTCTGAGAAATGCGCATTTACTCAACATATTTGACGCGCAATATCTCGTAGC 3180
Qy 3121 AACACTAAATTCTGAGAAATGCGCATTTACTCAACATATTTGACGCGCAATATCTCGTAGC 3180
Db 3181 GAAAAATACAGTAACCCCTTTAAATGACTATTTGATGTGATTTACGGGCTCGATTTTCG 3240
Qy 3181 GAAAAATACAGTAACCCCTTTAAATGACTATTTGATGTGATTTACGGGCTCGATTTTCG 3240
Db 3241 AAACGAATATATGCTCGAATTTGTGACCAACGAATTTAATTTGTGATTTTGTGTTCTT 3300
Qy 3241 AAACGAATATATGCTCGAATTTGTGACCAACGAATTTAATTTGTGATTTTGTGTTCTT 3300
Db 3301 TTGATATTTTGTGATCAATTAATAATTTTCCGTAAACAGACACACGCGCTACAGTACT 3360
Qy 3301 TTGATATTTTGTGATCAATTAATAATTTTCCGTAAACAGACACACGCGCTACAGTACT 3360
Db 3361 CTTTAAAGAGTTACAGTAGTTTTCGCTTCAAGATATTTTGAAGAATTTTAAACATTT 3420
Qy 3361 CTTTAAAGAGTTACAGTAGTTTTCGCTTCAAGATATTTTGAAGAATTTTAAACATTT 3420
Db 3421 TGAATAAAATATCATCTAATCATGTGCCAAAACGCTTTTTCAGAGTTTCGAGATTTTGA 3480
Qy 3421 TGAATAAAATATCATCTAATCATGTGCCAAAACGCTTTTTCAGAGTTTCGAGATTTTGA 3480
Db 3481 TTTTTCATTCACAGATATGCTTATTAACATATTAATTCATTAATGTGAATTTCTTG 3540
Qy 3481 TTTTTCATTCACAGATATGCTTATTAACATATTAATTCATTAATGTGAATTTCTTG 3540
Db 3541 TAGAAATTTTGGGCTTTTCTGTTCTAGTATGCTCTACTTTTGAATTTGCTCAACGAAAAA 3600
Qy 3541 TAGAAATTTTGGGCTTTTCTGTTCTAGTATGCTCTACTTTTGAATTTGCTCAACGAAAAA 3600
Db 3601 TCATGTGTTGTCAGAAATAAGTAAATAAGCGCATGCAATTTTATATATTTTCCCTAT 3660
Qy 3601 TCATGTGTTGTCAGAAATAAGTAAATAAGCGCATGCAATTTTATATATTTTCCCTAT 3660
Db 3661 TCATGTGTTGTCAGAAATAAGTAAATAAGCGCATGCAATTT 3700
Qy 3661 TCATGTGTTGTCAGAAATAAGTAAATAAGCGCATGCAATTT 3700

RESULT 2
LOCUS CEC48D1 39908 bp DNA INV 23-NOV-1998
DEFINITION Caenorhabditis elegans cosmid C48D1, complete sequence.
ACCESSION 281049
NID 91627677
VERSION 281049.1 GI:1627677
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 39908)
AUTHORS Burton,J.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwenematode.wustl.edu

LOCUS AQ191954 431 bp DNA GSS 03-NOV-1998
DEFINITION HS_3228_B2_F08_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3228 Col=16 Row=L, genomic survey
sequence.
ACCESSION AQ191954
NID 93590576
VERSION AQ191954.1 GI:3590576
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 431)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
TITLE Construction of a Characterized Clone Resource for Genomic
Sequencing
JOURNAL Unpublished (1998)
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3228 row: L column: 16
Class: BAC ends
High quality sequence stop: 431.
FEATURES
source
1..431
Location/Qualifiers
/organism="Homo sapiens"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
/db_xref="taxon:9606"
/clone="Plate=3228 Col=16 Row=L"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
BASE COUNT 186 a 70 c 59 g 116 t
ORIGIN
Query Match 2.6%; Score 26; DB 40; Length 431;
Best Local Similarity 86.1%; Pred. No. 1.96e-04;
Matches 31; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
DB 48 GAAAAAAAACTAATGCAAAATACAAAAAT 83
|||||
QY 4092 GAAAAAAAACTAATGCGAAATGAAAAAT 4127

Search completed: Sat Aug 7 19:20:22 1999
Job time : 1739 secs.


```

basal ganglia, thalamus, cerebellum, midbrain, pons and medulla."
/db_xref="GDB:3835515"
/db_xref="taxon:9606"
/clone="IMAGE:173480"
/clone_lib="Soares adult brain N2b5HB55Y"
/sex="Male"
/dev_stage="55-year old"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT      138 a      121 c      138 g      154 t       7 others
ORIGIN

Query Match          2.7%;   Score 27;   DB 31;   Length 558;
Best Local Similarity 80.4%;   Pred. No. 1.80e-05;
Matches 41; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Db      429 TGCTAAATGTATAATAGTATTAAACTTTNATGCATGTTCTAAT 479
        ||| ||||| | ||| | ||||| ||| | ||| ||||| |
Oy      4110 TCGGAATAATGAAAAAAT-GCATTAATAATACATTTTGTCATTCTTACAT 4159

RESULT    13
LOCUS     A1419399             194 bp      mRNA           EST           30-MAR-1999
DEFINITION ttf2b07.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2097493 3'
ACCESSION A1419399
VERSION   94265330
KEYWORDS  A1419399.1 GI:4265330
SOURCE    EST.
ORGANISM  human.
           Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 194)
AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project
JOURNAL   (CGAP/BTGP), Tumor Gene Index
COMMENT   Unpublished (1998)
On Mar 10, 1998 this sequence version replaced gi:2949049.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.lnl.gov/bdbr/image/image.html

Insert length: 2134 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 187.
Location/Qualifiers
1. 194
/organism="Homo sapiens"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) withaa modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCCGCATATCTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="IMAGE:2097493"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"

```

```

BASE COUNT      49 a      33 c      38 g      74 t      /lab_host="DH10B"
ORIGIN

Query Match
Best Local Similarity 93.3%; Score 26; DB 26; Length 194;
Matches 28; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 24 TTAATAATCCATTTTATTATTTCACACAT 53
      ||||| ||||| ||||| |||||
QY 3902 TTAATAATACATTTTGTATTTCACACAT 3931

RESULT 14
LOCUS AA9711176 260 bp mRNA EST 20-MAY-1998
DEFINITION op1f11.s1 NCI_CGAP_Kid6 Homo sapiens cDNA clone IMAGE:1575405 3'
ACCESSION AA9711176
NID 93146466
VERSION AA9711176.1 GI:3146466
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 260)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
On Jan 19, 1998 this sequence version replaced gi:2285245.

```

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 259.
Location/Qualifiers
1. 260

FEATURES
source

/organism="Homo sapiens"
/note="Organ: kidney; Vector: Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pooled kidney tumors. 5' adaptor sequence: 5'
GAATCGGCACGAG 3' 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.0 kb."
/db_xref="taxon:9606"
/clone="IMAGE:1575405"
/clone_lib="NCI_CGAP_Kid6"
/sex="mixed"
/tissue_type="kidney tumor"
/lab_host="SOLR (kanamycin resistant)"

BASE COUNT
74 a 32 c 38 g 112 t 4 others

ORIGIN

Query Match 2.6%; Score 26; DB 21; Length 260;
Best Local Similarity 72.2%; Pred. No. 1.96e-04;
Matches 39; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 200 TTCTTAATTTTCTTGGAATATNTTTGGGCGCTTCCAAAGNTTGTGTAAT 253
||||| | |||| |||| ||| | | | |||| ||| ||||| |||||
Cp 4420 TTCATATCCTTCATGGAATATGTATGTAGTTGGTCCAGAAGCTTGTGTAAT 4367

RESULT 15

TITLE	An Aspergillus nidulans EST Database
JOURNAL	Unpublished (1998)
COMMENT	On Aug 21, 1998 this sequence version replaced

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: T3
High quality sequence stop: 124.

FEATURES	Location/Qualifiers
source	1. .246

```

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
3' end of cDNA cloned into XhoI site of pBluescript"
/db_xref="taxon:5072"
/clone="w8a01a1"
/clone_lib="Aspergillus nidulans 24hr asexual
developmental and vegetative cDNA lambda zap library"
/tissue_type="vegetative mycelia, asexual structures"

```

Query Match	2.78;	Score 27;	DB 24;	Length 246;
Best Local	Similarity 80.0%;	Pred. No. 1.80e-05;		
Matches	36;	Conservative 0;	Mismatches 9;	Indels 0;
				Gaps 0;
Db	16	TTTTTTTTTTTTTTTAAACAACAAAAAATGAACCCCATGATTTTT	60	
Oy	3912	TTTTTTGATTTTTCACACATCACATGATTTAACCCCATTTATTTTT	3956	

RESULT	9				
LOCUS	N98032	345 bp	mRNA	EST	18-NOV-1996
DEFINITION	2086C3 czapFdd2.1, Debopam Chakrabarti Plasmodium falciparum cDNA clone PF2086C, mRNA sequence.				
ACCESSION	N98032				
NID	g1675066				
VERSION	N98032.1	GI:1675066			
KEYWORDS	EST.				
SOURCE	malaria parasite <i>P. falciparum</i> .				
ORGANISM	plasmodium falciparum				
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
AUTHORS	1 (bases 1 to 345)				
	Dame,J.B., Arnot,D.E., Bourke,P., Chakrabarti,D., Christodoulou,Z. Coppel,R., Cowman,A., Craig,A., Fischer,K., Foster,J., Goodman,N., Hinterberg,K., Holder,A.A., Holt,D., Kemp,D., Lanzer,M., Lim,A., Newbold,C., Ravetch,J.V., Reddy,G.R., Rubio,J., Schuster,S.M., Su,X.-Z., Thompson,J.K., Vital,F., Wellems,T.E. and Werner,E.				
TITLE	Current status of the Plasmodium falciparum genome project				
JOURNAL	Mol. Biochem. Parasitol. 79, 1-12 (1996)				
MEDLINE	97001675				
COMMENT	On Apr 14, 1993 this sequence version replaced gi:716382.				

Contact: Debopam Chakrabarti
 Department of Molecular Biology and Microbiology
 University of Central Florida
 Orlando, FL 32816-2360
 Tel: 407 384 2061
 Fax: 407 384 3095
 Email: dchak@pegasus.cc.ucf.edu
 Seq primer: T3.

FEATURES
 source
 Location/Qualifiers
 1..345
 /organism="Plasmodium falciparum"
 /strain="Dd2"

Query Match	2.7%;	Score 27;	DB 33;	Length 345;
Best Local Similarity	64.4%;	Pred. No. 1.80e-05;		
Matches	56;	Conservative	0;	Mismatches 31; Indels 0; Gaps 0;

Db	97	ATGTGAATAAAAAAAAAAATTAATTTTATTTTCATTTNACACATTTNATTTTGTATATAA	156
QY	3616	ATATGAATGACGAAAAATAGCAATTTTATATATATTTTCCCATTCATGTGTGCAGAA	3675
Db	157	TATAATTTAATTTCTTATACATTTT	183
QY	3676	AAATAGTAAAAAGCCGATGCATTTT	3702

BASE COUNT	137	a	32	c	16	g	158	t	2	others
ORIGIN										


```

/note="Vector: Lambda ZAP II; Site_1: EcoR I; Site_2: Xho
I; PolyA+ RNA, from asynchrononous blood stage parasites of
the Dd2 isolate cultured in vitro, was reverse transcribed
using an oligo dT-Xho I primer. Second strand was
prepared using RNase H and DNA polymerase I. EcoR I
adapters were ligated to the cDNA, and it was digested
with Xho I. Prepared fragments were ligated into EcoR I +
Xho I digested lambda ZAP II vector. "
/db_xref="taxon:5833"
/clone="PF2086C"
/clone_lib="czapFDD2.1, Debopam Chakrabarti"
/lab_host="E. coli XL-1 blue"

```

RESULT	10
LOCUS	AI262740 431 bp mRNA EST 29-NOV-1998
DEFINITION	qđ56el0.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:I733514 3' similar to contains Alu repetitive element;contains element XBR repetitive element ; mRNA sequence.
ACCESSION	AI262740
NID	93870943
VERSION	AI262740.1 GI:3870943
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 431) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
AUTHORS	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
TITLE	Unpublished (1997)
JOURNAL	On Apr 14, 1993 this sequence version replaced gi:692609.
COMMENT	

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 555 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 398.
 Location/Qualifiers
 1. .431

```

/organism="Homo sapiens"
/note="Vector: pT7T13D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5',
TGTTACCAATCTGAAGTGGAGCGGCCGCCCAATTTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I

```



```
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT      7 a      16 c      21 g      34 t      169 others
ORIGIN

Query Match      4.4%; Score 44; DB 17; Length 247;
Best Local Similarity 13.2%; Pred. No. 4.57e-26;
Matches 32; Conservative 11; Mismatches 95; Indels 4; Gaps 4;

Db 1 HMDCTMNTVWRGCCCBAMNMKHTHMTBBMCCVRRVGTITNNKGKNGRTTWNDCSDNA 60
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Cp 4499 CATTCACGAGGACTCGAGAGATTTC-TGTACATGCTTTCTCGCGAAAACAGCGCTTA 4441
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 61 HCRIVBMYIARSKYGYGTBYYSWNVDNTGGTGVGKTTVNVHSGWNNRCSNSVYVWBT 120
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Cp 4440 TGGTTGGTCATCGACAAGTTCATATCCTCTTCATGGAATATGTATTGAGTTGGTCCAG 4381
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 121 AYCDYBHYBDRANHVDTRCTNDRCYCNNTASDNGTSAT-KRVTGYDKTSDCGGCGWRK 179
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Cp 4380 AAGCTTGTGCTGAATGAGCAGATTACACTGATCCGAGAGAGACATCCGGTGA 4321
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 180 VTYGSSBYBRGVNVMVRTSMWTDKSTKMSMDMSRRSRVHYGRMBNK-KRGMSRNMWT 238
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Cp 4320 ATGAAGAGTGGCGGATGCTGAA-ATGTGTTGGTTTACGATCGATTGATACCCCAAT 4262
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 239 DT 240
   :|
Cp 4261 TT 4260

RESULT 4
LOCUS      AA754458      247 bp      mRNA      EST      20-JAN-1998
DEFINITION 97SN1784 Rice Immature Seed Lambda ZAP11 CDNA Library Oryza sativa
ACCESSION  AA754458
            CDNA clone 97SN1784, mRNA sequence.
NID        92801164
VERSION    AA754458.1 GI:2801164
KEYWORDS   EST.
SOURCE     Oryza sativa.
ORGANISM   Oryza sativa.

REFERENCE 1 (bases 1 to 247)
AUTHORS  Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
            Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
            Unpublished (1998)
            On Jan 14, 1998 this sequence version replaced gi:1797455.

JOURNAL
TITLE
COMMENT

FEATURES
source
    Contact: Eun M.Y.
    Department of Cyto genetics
    National Inst. of Agri. Sci. and Tech, RDA
    Suwon, Kyungido, Korea
    Tel: 82 331 290 0301
    Fax: 82 331 290 0307
    Email: myunesun20@sti.re.kr
    Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
    University, Yongin, Korea. 449-728 bnhahm@bioserver.myongji.ac.kr
    Seq primer: M13 Reverse Primer.
    Location/Qualifiers
        1..247
            /organism="Oryza sativa"
            /cultivar="Milyang23"
            /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
            XhoI; Directional cDNA library inserted into lambda ZAP11
            vector at 5' end with EcoRI and 3' end with Xho I site."
            /db_xref="taxon:4530"
            /map="6"
            /clone="97SN1784"
            /clone_lib="Rice Immature Seed Lambda ZAP11 CDNA Library"
```

```
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
BASE COUNT      7 a      16 c      21 g      34 t      169 others
ORIGIN

Query Match      3.7%; Score 37; DB 17; Length 247;
Best Local Similarity 12.9%; Pred. No. 4.84e-17;
Matches 27; Conservative 95; Mismatches 84; Indels 3; Gaps 3;

Db 20 MNKHTHMTBMBCCVRRVGTITNNKGKNGRTTWNDCSDNAHCRIVTBWYAR-SKYG 78
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 3731 ACTTCACAGCTGAAGACGAGAGACGCGGAGAAATACACACATCTTCTGCGCTCTCG 3790
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 79 TBYYSWNVDNTGGTGVGKTTVNVHSGWNNRCSNSVYVWBTAYCDY-BHYBDRANHVD 137
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 3791 TCTTCAGCATGTGAATGGATCTCGTCGATGTAAATAATGTCGAATATGTAAAAA 3850
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 138 TRCTNDRCYCNNTASDNGTSATKR-VTGYDKTSDCGGCGWRKVTYGSBYBRGVNMY 196
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 3851 TGCATGCGCTTTTTCACACTTCTGACACAATGAATGAGGGAATGTATTAATAAC 3910
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 197 RTTSMWTDKSTKMSMDMSRRSRVHYGRW 225
   :||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 3911 ATTTTGTATTTTTCACACATCATGAT 3939

RESULT 5
LOCUS      AF034173      2275 bp      mRNA      EST      30-MAR-1998
DEFINITION AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens CDNA
            clone ntcon2 contig, mRNA sequence.
ACCESSION  AF034173
            AF034173
NID        92707735
VERSION    AF034173.1 GI:2707735
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2275)
AUTHORS  Tripodis,N. and Ragoussis,J.
            Generation of a transcription map in the region immediately
            centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
            Unpublished (1997)
            On Jan 19, 1998 this sequence version replaced gi:2045115.

JOURNAL
TITLE
COMMENT

FEATURES
source
    Contact: Tripodis, Nikos
    Division of Medical and Molecular Genetics
    Guys Hospital
    7th floor, Guy's Tower, London SE1 9RT, UK
    Email: nikos@nki.nl.
    Location/Qualifiers
        1..2275
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="6p21.3"
            /clone="ntcon2 contig"
            /clone_lib="Human mRNA (Tripodis and Ragoussis)"
            /clone_lib="Human mRNA (Tripodis and Ragoussis)"

BASE COUNT      438 a      619 c      470 g      599 t      149 others
ORIGIN

Query Match      3.7%; Score 37; DB 20; Length 2275;
Best Local Similarity 12.4%; Pred. No. 4.84e-17;
Matches 17; Conservative 75; Mismatches 43; Indels 2; Gaps 2;

Db 1491 RKRMTGMYKMYRAMMMAMACMWYWKMRGKCKWKYKRYKKTSTYKSRMYW 1550
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 3668 GTGCAGAAAAATAGTAAAAAAGCGCATGATTTTTCGACAT-TTTTACATGGAACGACA 3726
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1551 YTTYTYWCCTSMKSASGAMMRWGYGSRSSRSRYWGYGSMGCGMYKRYRYSWT 1610
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 3727 GCTCACTTCACATGTGAAGACGAGACGCGG-AGAAATACACACACATCTTCTGCGTC 3785
```


RESULT 14
ID Q79970 standard; CDNA; 2485 BP.
AC Q79970;
DT 13-SEP-1995 (first entry)
DE Interleukin-1 beta converting enzyme homolog ced3 cDNA.
KW Interleukin-1 beta converting enzyme; homolog ced3;
KW oncogene bcl-2; programmed cell death; cancer treatment; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT cds 17..1528
FT /*tag= a
PN W09500160-A.
PD 05-JAN-1995.
PE 10-JUN-1994; U06630.
PR 24-JUN-1993; US-080850.
PA (GEHO) GEN HOSPITAL CORP.
PI Miura M, Yuan J;
DR WPI; 95-051742/07.
DR P-PSDB; R66770.
PT Promoting or preventing programmed cell death in vertebrate cells
PT - by inhibiting the activity of interleukin-1 beta converting
PT enzyme.
PS Example 1; Fig 2C; 116pp; English.
CC Q79970 encodes R66770 interleukin-1 beta converting enzyme homolog
CC ced3, increasing ced3 enzymatic activity can promote the
CC programmed cell death of cancer cells (pref. those overexpressing
CC the bcl-2 oncogene), this can be used as the basis of a new cancer
CC treatment. Alternatively by reducing ced3 enzymatic activity
CC programmed cell death can be inhibited, this may be useful in the
CC development of new cell lines which remain viable in culture for
CC extended or indefinite periods, independent of growth factors.
SQ Sequence 2485 BP; 684 A; 564 C; 480 G; 757 T;

Query Match 29.8%; Score 298; DB 14; Length 2485;
Best Local Similarity 100.0%; Pred. No. 2.66e-159;
Matches 298; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 552 catccgcaactcttattcaccgagatgctctctcggtacagttcaagtcgtaac 611
|||
Q 4303 CATCCGCCACTCTTCATTCACCGGATGCTCTCTCGGATACAGTTCAGTCTGTAATC 4362
Db 612 gctcattcagcaagcttctgacccaactcaatacatattcatgaagaagatgtaact 671
|||
QY 4363 GCTCATTCAGCAAGCTTCTGACCACTCATATATTCATGAAGAGATATGAAC 4422
Db 672 ttgtcgatgacccaacataagccgtgttttcgacgagaagaacatgtacagaacttct 731
|||
QY 4423 TTGTCGATGACCAACCACTAAGCCGTGTTTCGACGAGAAACCATGTACAGAACTTCT 4482
Db 732 cgaagtcctcgtagaatgtgcctcatcataaataatgaacactttgagcagatgccaacac 791
|||
QY 4483 CGAGTCCTCGTGAATGTGCTCATATATAATGAACACTTTGAGCAGATGCCAACAC 4542
Db 792 ggaatggtacccaagccgacaagacaactcttacaattgttcagatgcatggcta 849
|||
QY 4543 GGAATGGTACCAAGCCGACAAGACAATCTTACCAATTGTTCAGATGCATGGGCTA 4600

RESULT 15
ID Q64738 standard; DNA; 7653 BP.
AC Q64738;
DT 23-JUN-1994 (first entry)
DE ced-3 (C5940T) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT /*tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t repeat at 1490-1614"

FT repeat_unit 1490..1614
FT /*tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT /*tag= c
FT /number= Exon_1
FT 2367..2429
FT /*tag= d
FT /number= Intron_1
FT 2450..2575
FT /*tag= e
FT /number= Exon_2
FT 2576..2853
FT /*tag= f
FT /number= Intron_2
FT 2854..3107
FT /*tag= g
FT /number= Exon_3
FT 3108..4302
FT /*tag= h
FT /number= Intron_3
FT 3126..3243
FT /*tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT 3329..3396
FT /*tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT 3487..3759
FT /*tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT 3782..4070
FT /*tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT 4303..4634
FT /*tag= m
FT /number= Exon_4
FT 4635..5546
FT /*tag= n
FT /number= Intron_4
FT 4688..4719
FT /*tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5221..5330
FT /*tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5547..5760
FT /*tag= q
FT /number= Exon_5
FT 5761..5814
FT /*tag= r
FT /number= Intron_5
FT 5815..5942
FT /*tag= s
FT /number= Exon_6
FT 5943..6297
FT /*tag= t
FT /note= "C>T, from allele n1165"
FT 6298..6537
FT /*tag= v
FT /rpt_type= INVERTED
FT /note= "w"
FT /number= Exon_7

FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT exon 5547..5760
FT /tag= q
FT /number= Exon_5
FT mutation 5757
FT /tag= r
FT /note= "G>A, from allele n2433"
FT intron 5761..5814
FT /tag= s
FT /number= Intron_5
FT exon 5815..5942
FT /tag= t
FT /number= Exon_6
FT intron 5943..6297
FT /tag= u
FT /number= Intron_6
FT repeat_region 6062..6138
FT /tag= v
FT /rpt_type= INVERTED
FT exon 6298..6537
FT /tag= w
FT /number= Exon_7
FT intron 6538..7012
FT /tag= x
FT /number= Intron_7
FT repeat_unit 6567..6625
FT /tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT repeat_unit 6905..6965
FT /tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT exon 7012..7075
FT /tag= aa
FT /number= Exon_8
FT
FT WO9325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-8977788.
PR 20-NOV-1992; US-979638.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.
DR P-PSDB; R53281.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127P; English.
CC The sequences given in Q64735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

Query Match 99.8%; Score 998; DB 9; Length 7653;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 3601 tcatgtggttctcatatgaatgacccaaaatagcaatttttatatatattccctat 3660
QY 3601 TCATGTGTTGTTTCATATGAATGACGAAAAATAGCAATTTTATATATTTTCCCTAT 3660
Db 3661 tcatgtgtgcagaaaaatagtaaaaaagcgcatgcattttcgacatttttacatcga 3720
QY 3661 TCATGTTGTGCAGAAAAATAGTAAAAAAGCGCATGCAATTTTTCGACATTTTACATCGA 3720
Db 3721 acgacagctcacttcacatgctgaagcagagagacgcgagagaataaccacacatcttct 3780
QY 3721 ACGACAGCTCACTTCACATGCTGAAGACGAGAGACGCGGAGAAATACACACATCTTCT 3780
Db 3781 gcgtctctcgctctcagcatgtgaaatggtatctcggtcgatgtaaaaaatgtcgaata 3840
QY 3781 GCGTCTCTCGTCTTCAGCATGTGAATGGGATCTCGGTGATGTAAAAAATGTCGAATA 3840
Db 3841 atgtaaaaaatgcatgcgttttttacaactttctgcacaaatgaaataggggaaatgt 3900
QY 3841 ATGTAAAAAATGCATGCGTTTTTTTACACTTTTCTGCACAAATGAATAGGGGAAAAATGT 3900
Db 3901 attaaaatacatttttltglattttcaacatcacatgataaaccatatttttcgt 3960
QY 3901 ATTAAATACATTTTTTGTATTTTTCACACATCACATGATTAACCCCATTTATTTTCGT 3960
Db 3961 gagcaactlaaaaaagtagaataatagagcgaaaccacaaatttctcaagatatacc 4020
QY 3961 GAGCAACTTAAAAAGTAGAGAATATTAGAGCGAAACCAAAATTTCTTCAAGATATACC 4020
Db 4021 ttattgtaaatatagatgtttaataagcatatcttgaatgaagtacgaaaaatatgt 4080
QY 4021 TTTATGTATAATTATAGATGTTAATAGCATATCTTGAATGAAGTCAGCAAAAAATATGT 4080
Db 4081 gcgaacacacctgaaaaaaatcacaattctgcgaaaaattgaaaaaaatgcatlaaaataca 4140
QY 4081 GCGAAACACCTGAAAAAATCAAAAAATTTCTGCGAAAAATTGAAAAATGCATTAAAAATACA 4140
Db 4141 ttttgcatttttctacatcacatgaaatgtagaaaatlaaaaggaaatcaaaatttcta 4200
QY 4141 TTTTGCATTTTCTACATCACATGAATGTAGAAAAATTAAAGGGAATCAAAATTTCTA 4200
Db 4201 gaggatataattgaaatgaaacatgycgaaatlaaaatgtgcgaaacgtcaaaaaagaga 4260
QY 4201 GAGGATATAATTGAATGAACATTCGGAATTTAAATGTGCGAAACGTCAAAAAAGAGA 4260
Db 4261 aattggtatcaaaaatcgatccctaaaccacacacatttcagcatccgcacactctcat 4320
QY 4261 AATTGGGTATCAAAAATCGATCCCTAAACCACACATTTTCAGCATCCGCCAATCTTCAT 4320
Db 4321 tcaccggtatcttctctcgcgatacagttcaagtgcgtaatcgctcattcagcaaaagct 4380
QY 4321 TCACCGGATGCTCTTCTCGGATACAGTTCAGTCCGTAATCGCTCATTCAGCAAACTT 4380
Db 4381 ctggaaccaactcaatacatattccatgaagagatgaactttgtcgatgcaccaacca 4440
QY 4381 CTGGAACCAACTCAATATACATATTCATGAAGAGGATATGAACCTTTGTGATGCACCAACCA 4440
Db 4441 taagccgtgttttcgacgagaaaaaccatgtacagaaacttcgagtcctcgtgaaatgt 4500
QY 4441 TAAGCCGTGTTTTCGACGAGAAAAACCATGTACAGAAACTTCTCGAGTCCCTGTAATGT 4500
Db 4501 gcctcatcataaataatgaacacttgagcagatgcacaacggaatggtaccaaagccg 4560
QY 4501 GCCTCATCATTAATAATGAACACTTTGAGCAGATGCCAACACGGAATGTACCAAGCCG 4560
Db 4561 acaagacaatcttacaatttgttcagatgcatggcta 4600
QY 4561 ACAAGACAATCTTACCATTTGTTCAGATGCATGGGCTA 4600

CC splicing mutations (see also Q64735-45). These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2429 A; 1453 C; 1272 G; 2499 T;

Query Match 99.8%; Score 998; DB 9; Length 7653;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 3601 tcatgtgtttgttcatatgaatgacccaaaaatagcaatttttatatatattccctat 3660
|||||
QY 3601 TCATGTGTTGTTTCATATGAATGACGAAAAATAGCAATTTTATATATTTTCCCTAT 3660
Db 3661 tcatgttgcagaaaaaatagtaaaaaagcgcatgcatlltcgacatttttacatcga 3720
|||||
QY 3661 TCATGTTGTGAGAAAAATAGTAAAAAGCGCATGTTTTCGACATTTTTCATCGA 3720
Db 3721 acgacagctcacttcacatgctgaagacgagagacgcgagaaataccacacatcttct 3780
|||||
QY 3721 ACGACAGCTCAGTTCACATGCTGAAGACGAGAGACGCGGAGAAATACACACATCTTCT 3780
Db 3781 gcgtctcgtcttcacgatgtgaatggatctcgatgtaaaaaaatgtcgaata 3840
|||||
QY 3781 GCGTCTCGTCTTCAGCATGTGAATGGGATCTCGGTGATGTAAAAAATGTGGAATA 3840
Db 3841 atgtaaaaaatgcatgctllttttacactlltcgcacaaatagatagggggaaatgt 3900
|||||
QY 3841 ATGTAAAAAATGCATGCGTTTTTCACACTTTTTCGACAAATGATAGGGGAAATGT 3900
Db 3901 attaaaaacatttttgtattttcacaacacacatgattaaccccatatttttcgtt 3960
|||||
QY 3901 ATTAATAATACATTTTGTATTTTTCAACATCAGATGATTAACCCCATTTTTCGTT 3960
Db 3961 gagcaacttaaaagtagaataatagagcgaaaaaccaaatttctcaagatatacc 4020
|||||
QY 3961 GAGCAACTTAAAAAGTAGAAGATATTAGAGCGAAAAACCAAAATTTCTCAAGATATTACC 4020
Db 4021 ttatgtataatatagatgttaataagcatatcttgaatgaagtcagcaaaatatgt 4080
|||||
QY 4021 TTTATTGATAATTATAGATGTTAATAAGCATATCTTGAATGAAGTCAGCAAAAAATATGT 4080
Db 4081 gcgaacacactgaaaaaaatcaaaaattctgcgaaaaattgaaaaaatgcatlaaataca 4140
|||||
QY 4081 GCGAAACACCTGAAAAAATCAAAAATCTCGGAAAAATTGAAAAAATGCATTAATAATACA 4140
Db 4141 ttttgcatttttctacatcacatgtaatgttgaataaataaagggaatcaaaatttcta 4200
|||||
QY 4141 TTTTGCATTTTCTACATCAGATGATGTAGAAAAATTAAAGGGAATCAAAATTTCTA 4200
Db 4201 gaggatataatgaatgaataacattgcgaataataaaatgtgcgaacgtcaaaaaagaga 4260
|||||
QY 4201 GAGGATATAATTGAATGAACATTGCGAAATTAAATGTGCGAAACGTCAAAAAAGAGGA 4260
Db 4261 aattgggtatcaaaatcgatccataaaaccaaacaatttcagcatccgccaactctcat 4320
|||||
QY 4261 AATTGGGTATCAAAATCGATCCTAAAAACAACACATTTTCAGCATCCGCCAATCTTCAT 4320
Db 4321 tcaccggtatctctctcgatatacagttcaagtcgtaatcgctcatcagcaagcct 4380
|||||
QY 4321 TCACCGGATGCTCTCTCTCGGATACAGTTCAGTCAAGTCGTAATCGTCAATCAGCAAAAGCTT 4380
Db 4381 ctggaccaactcaatatacatattccatgaagagagatatgaactttgtcgatgcaccacca 4440
|||||
QY 4381 CTGGACCAACTCAATATACATATTCCATGAAGAGGATATGAACCTTGTGATGCACCAACCA 4440
Db 4441 taagccgtgttttcgacgagaaaaaacatgtacagaaaactctcgagtcctcgtggaatgt 4500
|||||
QY 4441 TAAGCCGTGTTTTCGACGAGAAAAACCATGTACAGAAACTTCTCGAGTCTCTGTGGAATGT 4500

Db 4501 gccctcatataataatgaacactttgacgagatgccaacaggaatggtaccagggccg 4560
|||||
QY 4501 GCCTCATCATATAATAATGAACACTTTGACAGATGCCAACACGGAATGTACCAAGGCCG 4560
Db 4561 acaaggaacatttaccattgttcagatgcatggcta 4600
|||||
QY 4561 ACAAGGACAACTTACCAATTGTTTCAGATGCATGGGCTA 4600

RESULT 13
ID Q64737 standard; DNA; 7653 BP.
AC Q64737;
DT 23-JUN-1994 (first entry)
DE ced-3 (G5757A) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT FT /*tag= a
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t repeat at 1490-1614"
FT FT 1490..1614
FT FT /*tag= b
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT FT 2232..2366
FT FT /*tag= c
FT FT /number= Exon_1
FT FT 2367..2429
FT FT /*tag= d
FT FT /number= Intron_1
FT FT 2450..2575
FT FT /*tag= e
FT FT /number= Exon_2
FT FT 2576..2853
FT FT /*tag= f
FT FT /number= Intron_2
FT FT 2854..3107
FT FT /*tag= g
FT FT /number= Exon_3
FT FT 3108..4302
FT FT /*tag= h
FT FT /number= Intron_3
FT FT 3126..3243
FT FT /*tag= i
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3339-3396"
FT FT 3329..3396
FT FT /*tag= j
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT FT 3487..3759
FT FT /*tag= k
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT FT 3782..4070
FT FT /*tag= l
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 3487-3759"
FT FT 4303..4634
FT FT /*tag= m
FT FT /number= Exon_4
FT FT 4635..5546
FT FT /*tag= n
FT FT /number= Intron_4
FT FT 4688..4719
FT FT /*tag= o
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT 5221..5330
FT FT repeat_unit

FT	/*tag= a	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t repeat at 1490-1614"	
FT	1490..1614	
FT	/*tag= b	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 1356-1472"	
FT	2232..2366	
FT	/*tag= c	/number= Exon_1
FT	2310	
FT	/*tag= d	/note= "C>T, from allele n1040"
FT	2367..2429	
FT	/*tag= e	/number= Intron_1
FT	2430..2575	
FT	/*tag= f	/number= Exon_2
FT	2487	
FT	/*tag= g	/note= "G>A, fromm allele n718"
FT	2576..2853	
FT	/*tag= h	/number= Intron_2
FT	2854..3107	
FT	/*tag= i	/number= Exon_3
FT	3108..4302	
FT	/*tag= j	/number= Intron_3
FT	3126..3243	
FT	/*tag= k	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 3329-3396"	
FT	3329..3396	
FT	/*tag= l	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 3126-3243"	
FT	3487..3759	
FT	/*tag= m	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 3782-4070"	
FT	3782..4070	
FT	/*tag= n	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat 3487-3759"	
FT	4303..4634	
FT	/*tag= o	/number= Exon_4
FT	4635..5546	
FT	/*tag= p	/number= Intron_4
FT	4688..4719	
FT	/*tag= q	/rpt_type= INVERTED
FT	/note= "Repeat 3"	
FT	5221..5330	
FT	/*tag= r	/rpt_type= INVERTED
FT	/note= "Repeat 3"	
FT	5547..5760	
FT	/*tag= s	/number= Exon_5
FT	5757	
FT	/*tag= t	/note= "G>A, from allele n2433"
FT	5761..5814	
FT	/*tag= u	/number= Intron_5
FT	5815..5942	
FT	/*tag= v	/number= Exon_6

FT	mutation	5940	/tag= w
FT		/note= "C>T, from allele n1165"	
FT	intron	5943..6297	
FT		/tag= x	
FT		/number= Intron_6	
FT	repeat_region	6062..6138	
FT		/tag= y	
FT		/rpt_type= INVERTED	
FT	mutation	6297	
FT		/tag= z	
FT		/note= "G>A, fromn allele n717"	
FT	exon	6298..6537	
FT		/tag= aa	
FT		/number= Exon_7	
FT	mutation	6322	
FT		/tag= ab	
FT		/note= "C>T, from allele n1949"	
FT	mutation	6372	
FT		/tag= ac	
FT		/note= "G>A, from allele n1286"	
FT	mutation	6434	
FT		/tag= ad	
FT		/note= "C>T, fromm alleles n1129 and n1164"	
FT	mutation	6485	
FT		/tag= ae	
FT		/note= "C>T, from allele n2430"	
FT	mutation	6535	
FT		/tag= af	
FT		/note= "G>A, from allele n2426"	
FT	intron	6538..7012	
FT		/tag= ag	
FT		/number= Intron_7	
FT	repeat_unit	6567..6625	
FT		/tag= ah	
FT		/rpt_type= INVERTED	
FT		/note= "Inverted w.r.t. repeat at 6905-6965"	
FT	repeat_unit	6905..6965	
FT		/tag= ai	
FT		/rpt_type= INVERTED	
FT		/note= "Inverted w.r.t. repeat 6567-6625"	
FT	exon	7013..7075	
FT		/tag= aj	
FT		/number= Exon_8	
FT	mutation	7020	
FT		/tag= ak	
FT		/note= "C>T, from allele n1163"	
PN	WO9325685-A.		
PD	23-DEC-1993.		
PF	14-JUN-1993;	U05701.	
PR	12-JUN-1992;	US-897788.	
PR	20-NOV-1992;	US-979638.	
PA	(MAST) MASSACHUSETTS INST TECHNOLOGY.		
PI	Horvitz HR, Shaham S, Yuan J;		
DR	WPI; 94-007542/01.		
DR	P-PSDB; R47466.		
PT	Isolated C elegans cell death genes ced-3 and ced-4 - used to		
PT	develop agents to increase or prevent cell death in organisms		
PS	Claim 2; Fig 4; 127pp; English.		
CC	This sequence represents the C. elegans ced-3 gene. A 2.8 kb mRNA		
CC	was identified as the ced-3 transcript and was most abundant in		
CC	embryos, but was also detected in larvae and young adults, suggesting		
CC	that ced-3 is expressed not only in cells undergoing cell death. The		
CC	four largest introns as well as sequences 5' of the start codon		
CC	contain repetitive elements, some of which have been characterised		
CC	in non-coding regions of other C. elegans genes, such as fem-1, lin-12		
CC	and myoD. The Ced-3 protein is 503 amino acids in length. Ced-3 is		
CC	highly hydrophilic with no significant hydrophobic region that might		
CC	be a transmembrane region. One region of Ced-3 is very rich in serine.		
CC	It is thought that this region is involved in protein-protein inter-		
CC	actions, similar to acid blobs in the ced-3 gene, eight of the mutations		
CC	are missense mutations, two are nonsense mutations and two are putative		

FT	intron	5943..6297
FT		/*tag= t
FT		/number= Intron_6
FT	repeat_region	6062..6138
FT		/*tag= u
FT		/rpt_type= INVERTED
FT	exon	6298..6537
FT		/*tag= v
FT		/number= Exon_7
FT	mutation	6322
FT		/*tag= w
FT		/note= "C>T, from allele n1949"
FT	intron	6538..7012
FT		/*tag= x
FT		/number= Intron_7
FT	repeat_unit	6567..6625
FT		/*tag= y
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 6905-6965"
FT	repeat_unit	6905..6965
FT		/*tag= z
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat 6567-6625"
FT	exon	7012..7075
FT		/*tag= aa
FT		/number= Exon_8
PN	WO9325685-A.	
PD	23-DEC-1993.	
PF	14-JUN-1993;	U05701.
PR	12-JUN-1992;	US-897788.
PR	20-NOV-1992;	US-979638.
PA	(MASI) MASSACHUSETTS INST TECHNOLOGY.	
PI	Horvitz HR, Shaham S, Yuan J;	
DR	WPI; 94-007542/01.	
DR	P-PSDB; R53283.	
PT	Isolated C elegans cell death genes ced-3 and ced-4 - used to	
PT	develop agents to increase or prevent cell death in organisms	
PS	Claim 14; Fig 4: 127pp; English.	
CC	The sequences given in Q64735-45 represent mutations of the C. elegans	
CC	ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and	
CC	was most abundant in embryos, but was also detected in larvae and young	
CC	adults, suggesting that ced-3 is expressed not only in cells undergoing	
CC	cell death. The four largest introns as well as sequences 5' of the	
CC	start codon contain repetitive elements, some of which have been	
CC	characterised in non-coding regions of other C. elegans genes, such	
CC	'as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in	
CC	length. Ced-3 is highly hydrophilic with no significant hydrophobic	
CC	region that might be a transmembrane region. One region of Ced-3 is	
CC	very rich in serine. It is thought that this region is involved in	
CC	protein-protein interactions, similar to acid blobs in transcription	
CC	factors. Of the mutations which occur within the ced-3 gene, eight of	
CC	the mutations are missense mutations, two are nonsense mutations and	
CC	two are putative splicing mutations. These mutations establish the	
CC	null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,	
CC	function is not essential for viability. The ced-3 and ced-4 gene	
CC	products may be used to develop agents for treating conditions	
CC	characterised by cell deaths, such as myocardial infarction, stroke,	
CC	degenerative disease, traumatic brain injury, hypoxia, pathogenic	
CC	infection, aging or hair loss.	
CC	Sequence 7653 BP;	2429 A; 1452 C; 1272 G; 2500 T;

Query Match	99.88;	Score 998;	DB 9;	Length 7653;
Best Local Similarity	99.98;	Pred. No. 0.00e+00;		
Matches 999;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

Accession	Sequence	Length
Dd 3601	tcatgtgttctgttcataatgaatgacccaataattttatatatttccctat	3660
Qy 3601	TCATGTGTTTGTTCATATGAATGACGAAAAATGCAATTTTATATTTTCCCTAT	3660
Dd 3661	tcatgttctgtcagaanaaatagtaaaaagcgcatttctgcacattttacatcga	3720
Qy 3661	TCATGTGTGCGAAGAAATAGTAAAAAAGCCGATTTTTCGACATTTTTCACATCGA	3720

Db	3721	acgacagctcacttcacatgctgaaagcagagacgcgagaaataaccacacttctc	3780
QY	3721	acgacagctcacttcacatgctgaaagcagagacgcgagaaataaccacacttctc	3780
Db	3781	gcgtctctcgctcttcacgatgtgaaatgagatctcggtcgatgtaaaaaaatgctgaata	3840
QY	3781	gcgtctctcgctcttcacgatgtgaaatgagatctcggtcgatgtaaaaaaatgctgaata	3840
Db	3841	atgtaaaaaatgcatgcgtttttttacactttctgcacaaatgtaatagggggaaaaatgt	3900
QY	3841	atgtaaaaaatgcatgcgtttttttacactttctgcacaaatgtaatagggggaaaaatgt	3900
Db	3901	atataaatacatcttttttgtatttttccaacatcacatgatattaaccccatlattttcggt	3960
QY	3901	atataaatacatcttttttgtatttttccaacatcacatgatattaaccccatlattttcggt	3960
Db	3961	gagcaacttaaaaagtagagaatatatagagcgaaaaaaccaaaattcttcagaatatacc	4020
QY	3961	gagcaacttaaaaagtagagaatatatagagcgaaaaaaccaaaattcttcagaatatacc	4020
Db	4021	tttatgtataatlatagatgtttaataagcatatcttgaatgaaagtccagcaaaaatatgt	4080
QY	4021	tttatgtataatlatagatgtttaataagcatatcttgaatgaaagtccagcaaaaatatgt	4080
Db	4081	gcgaacacacctgnaaaaaatcaaaaattctgcgnaaattgnaaaaatgcattaaaatata	4140
QY	4081	gcgaacacacctgnaaaaaatcaaaaattctgcgnaaattgnaaaaatgcattaaaatata	4140
Db	4141	tttttgcatttttctcatcacatgtaatgtgaaaaatltaaaagggaaatccaanaattcta	4200
QY	4141	tttttgcatttttctcatcacatgtaatgtgaaaaatltaaaagggaaatccaanaattcta	4200
Db	4201	gaggatataatlgnaatgnaaacatttgcgnaaatcaaaaatgttgcgnaaacgctcaaaaaagagga	4260
QY	4201	gaggatataatlgnaatgnaaacatttgcgnaaatcaaaaatgttgcgnaaacgctcaaaaaagagga	4260
Db	4261	aatttgygtatcaaaaatcgatccctaanaaccacacacatttcagcatccgccaactcttcat	4320
QY	4261	aatttgygtatcaaaaatcgatccctaanaaccacacacatttcagcatccgccaactcttcat	4320
Db	4321	tcaccggatgctctctctcctcgatatcaggttcaagtctgtaatcgctcatctcagcaaaagctt	4380
QY	4321	tcaccggatgctctctctcctcgatatcaggttcaagtctgtaatcgctcatctcagcaaaagctt	4380
Db	4381	ctggaaccaactcaatatacatatccatggaagagatatgnaactttgtcgatgcaaccaacca	4440
QY	4381	ctggaaccaactcaatatacatatccatggaagagatatgnaactttgtcgatgcaaccaacca	4440
Db	4441	taagccggtgttttcgacgagaaaaaccatgtacagaaaacttctcgagtccctcgtgtaatgt	4500
QY	4441	taagccggtgttttcgacgagaaaaaccatgtacagaaaacttctcgagtccctcgtgtaatgt	4500
Db	4501	gcctcatcataataataatgaaacactttgagcgagatgccaacacggaatggttaaccaagcg	4560
QY	4501	gcctcatcataataataatgaaacactttgagcgagatgccaacacggaatggttaaccaagcg	4560
Db	4561	acaagggacaatcttaaccaatttgttcagatgcatggygcta 4600	
QY	4561	acaagggacaatcttaaccaatttgttcagatgcatggygcta 4600	

RESULT	12
ID	Q54666 standard; DNA; 7653 BP.
AC	Q54666;
DT	23-JUN-1994 (first entry)
DE	ced-3 gene.
KW	C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW	embryogenesis; cell death; hydrophilic; transmembrane; region;
KW	hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW	protein synthesis; null phenotype; calcium-binding domain; ss.
OS	Caenorhabditis elegans.
FN	Key
FT	repeat_unit location/Qualifiers 1356..1472

Db 3601 tcatgtgttcttccatatgaatgacccaaaaatagcaatttttatataatttccctat 3660
|||||
QY 3601 TCATGTGTTGTTTCATATGAATGACGAAAAATAGCAATTTTATATATTTTCCCTAT 3660
Db 3661 tcatgtgtgcagaaaaatagtaaaagcgcgtcatttccgacatttccatcga 3720
|||||
QY 3661 TCATGTTGTGAGAAAAATAGTAAAAAGCGCATGATTTTTCGACATTTTTCATCGA 3720
Db 3721 acgacagctcacttcacatgctgaagacgagagcgcgagaataaccacatcttct 3780
|||||
QY 3721 ACGACAGCTCAGCTTCACATGCTGAAGACGAGACGCGGAGAAATACACACATCTTCT 3780
Db 3781 gcgtctctcgtcttcagcatgtgaatgagatcgtcgtcgatgttaaaaaatgtcgata 3840
|||||
QY 3781 GCGTCTCTCGTCTTCAGCATGTGAATGGGATCTCGGTCGATGTAAAAAATGTCGATA 3840
Db 3841 atgtaaaaaatgcatggtttttttacacatttctgcacaaatgaatagggggaatgt 3900
|||||
QY 3841 ATGTAAAAAATGCATGCGTTTTTCACACTTTCTGCACAAATGAATAGCGGGAATGT 3900
Db 3901 attaaaaatacatttttctgatttttccaacatcacatgattaaacccattatttctgt 3960
|||||
QY 3901 ATTTAAATACATTTTGTATTGTTTTCACATCATGATTAACCCCATTTTTCGTT 3960
Db 3961 gagcaacttaaaaaagtagaataatagagcgaaaaaccaaattctcagaatatacc 4020
|||||
QY 3961 GAGCAACTTAAAAAGTAGAATAATTAGAGCGAAAAACCAAAATTTCTCAAGATATTACC 4020
Db 4021 ttattgataattatagtgtaataagcatatcttgaatgaatgaatgcagcaaaatatgt 4080
|||||
QY 4021 TTTATTGATTAATTATGATGTTAATAAGCATATCTTGAATGAAGTCAAGCAAAATATGT 4080
Db 4081 gcgaaacacctgaaaaaaatcaaaaattctgcgaaaaatgaaaaaatgataataata 4140
|||||
QY 4081 GCGAAACACCTGAAAAAATCAAAATTTCTGCGAAAAATTGAAAAAATGCTTAAATACA 4140
Db 4141 ttttgcatttttctacatcacatgtaatgtagaanaattaaagggaatcaaaattctta 4200
|||||
QY 4141 TTTTGCATTTTCTACATCACATGAATGTAGAAAAATTTAAAGGGAATCAAAATTTCTTA 4200
Db 4201 gaggatataatgaatgaataacattgcgaataataaagtgcgaacgtaaaaaagaga 4260
|||||
QY 4201 GAGGATATAATTGAATGAACAATTGCGAAATTAAAAATGTGCGAAACGTCAAAAAGAGA 4260
Db 4261 aattgggtatcaaaatcgatcctaataacacacatttcagcatccgccaactctcat 4320
|||||
QY 4261 AATTGGGTATCAAAATCGATCCTAAAAACCAACATTTTCAGCATCCGCCAATCTTCAT 4320
Db 4321 tcaccggtgctctctcgcgatacagttcaagtcgtaatcgctcatcagcaaacgt 4380
|||||
QY 4321 TCACCGGATGCTCTCTCGGATACAGTTCAGTCAAGTCGTAATCGCTCATTCAGCAAAAGCTT 4380
Db 4381 ctggaccaactcaatacatattccatgaagagatatgaactttgtcgatgcaccaacca 4440
|||||
QY 4381 CTGGACCAACTCAATACATATTCATGAAGAGGATATGAACCTTGTGATGCACCAACCA 4440
Db 4441 taagccgtgttttcgacgagaaaaaacattgtacagaacttctcgagtcctcgttgaatgt 4500
|||||
QY 4441 TAAAGCGTGTTCGACGAGAAAAACCATGTACAGAAACTTCTCGAGTCTCTGTGAATGT 4500
Db 4501 gccctcatcataataataaacacatttgagcagatgccaaacacggaatggtaccaaagcg 4560
|||||
QY 4501 GCCTCATCATAAATAATGACACTTTGAGCAGATGCCAACACGGAATGTTACCAAGGCGG 4560
Db 4561 acaaggacaatcttaccattgttcagatgcattgggcta 4600
|||||
QY 4561 ACAAGGACAATCTTACCATTGTTGTCAGATGCATGGGCTTA 4600

RESULT 11
ID 064740 standard; DNA; 7653 BP.
AC 064740;

DT 23-JUN-1994 (first entry)
DE ced-3 (C6322T) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key
FT repeat_unit
FT 1356..1472
FT /tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1490-1614"
FT 1490..1614
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT 2232..2366
FT /tag= c
FT /number= Exon_1
FT 2367..2429
FT /tag= d
FT /number= Intron_1
FT 2450..2575
FT /tag= e
FT /number= Exon_2
FT 2576..2853
FT /tag= f
FT /number= Intron_2
FT 2854..3107
FT /tag= g
FT /number= Exon_3
FT 3108..4302
FT /tag= h
FT /number= Intron_3
FT 3126..3243
FT /tag= i
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT 3329..3396
FT /tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT 3487..3759
FT /tag= k
FT /number= Exon_4
FT 4635..5546
FT /tag= n
FT /number= Intron_4
FT 4688..4719
FT /tag= o
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5221..5330
FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5547..5760
FT /tag= q
FT /number= Exon_5
FT 5761..5814
FT /tag= r
FT /number= Intron_5
FT 5815..5942
FT /tag= s
FT /number= Exon_6


```
Db      4561  accaaggacaatcttaccgaattgttcagatgcattggccta 4600
          |||
OY      4561  ACAAGGACATCTTACCAATTGTTCAGATGCATGGGCTA 4600

RESULT  10
ID      Q64743 standard; DNA; 7653 BP.
AC      Q64743;
DT      23-JUN-1994 (first entry)
DE      ced-3 (C6485T) gene.
KW      C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW      embryogenesis; cell death; hydrophilic; transmembrane; region;
KW      hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW      protein synthesis; null phenotype; calcium-binding domain; ss.
OS      Caenorhabditis elegans.
FH      Key
FT      repeat_unit
          Location/Qualifiers
          /tag= a
          /rpt_type= INVERTED
          /note= "Inverted w.r.t repeat at 1490-1614"
FT      repeat_unit
          /tag= b
          /rpt_type= INVERTED
          /note= "Inverted w.r.t. repeat at 1356-1472"
FT      exon
          /tag= c
          /number= Exon_1
          /tag= d
          /number= Intron_1
          /tag= e
          /number= Exon_2
          /tag= f
          /number= Intron_2
          /tag= g
          /number= Exon_3
          /tag= h
          /number= Intron_3
          /tag= i
          /rpt_type= INVERTED
          /note= "Inverted w.r.t. repeat at 3329-3396"
FT      repeat_unit
          /tag= j
          /rpt_type= INVERTED
          /note= "Inverted w.r.t. repeat at 3126-3243"
FT      repeat_unit
          /tag= k
          /rpt_type= INVERTED
          /note= "Inverted w.r.t. repeat at 3782-4070"
FT      repeat_unit
          /tag= l
          /rpt_type= INVERTED
          /note= "Inverted w.r.t. repeat 3487-3759"
FT      exon
          /tag= m
          /number= Exon_4
          /tag= n
          /number= Intron_4
          /tag= o
          /rpt_type= INVERTED
          /note= "Repeat 3"
FT      repeat_unit
          /tag= p
          /rpt_type= INVERTED
          /note= "Repeat 3"
```

```
FT      exon
          /tag= q
          /number= Exon_5
          /tag= r
          /number= Intron_5
          /tag= s
          /number= Exon_6
          /tag= t
          /number= Intron_6
          /tag= u
          /rpt_type= INVERTED
          /note= "C>T, from allele n2430"
FT      intron
          /tag= v
          /number= Exon_7
          /tag= w
          /note= "C>T, from allele n2430"
FT      intron
          /tag= x
          /number= Intron_7
          /tag= y
          /rpt_type= INVERTED
          /note= "Inverted w.r.t. repeat at 6905-6965"
FT      repeat_unit
          /tag= z
          /rpt_type= INVERTED
          /note= "Inverted w.r.t. repeat 6567-6625"
FT      exon
          /tag= aa
          /number= Exon_8

WO9325685-A.
PD      23-DEC-1993.
PF      14-JUN-1993; U05701.
PR      12-JUN-1992; US-897788.
PR      20-NOV-1992; US-979638.
PA      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI      Horvitz HR, Shahan S, Yuan J;
DR      WPI; 94-007542/01.
DR      P-PSDB; R53286.
PT      Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT      develop agents to increase or prevent cell death in organisms
PS      Claim 14; Fig 4; 127pp; English.
CC      The sequences given in Q64735-45 represent mutations of the C. elegans
CC      ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC      was most abundant in embryos, but was also detected in larvae and young
CC      adults, suggesting that ced-3 is expressed not only in cells undergoing
CC      cell death. The four largest introns as well as sequences 5' of the
CC      start codon contain repetitive elements, some of which have been
CC      characterised in non-coding regions of other C. elegans genes, such
CC      as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC      length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC      region that might be a transmembrane region. One region of Ced-3 is
CC      very rich in serine. It is thought that this region is involved in
CC      protein-protein interactions, similar to acid blobs in transcription
CC      factors. Of the mutations which occur within the ced-3 gene, eight of
CC      the mutations are missense mutations, two are nonsense mutations and
CC      two are putative splicing mutations. These mutations establish the
CC      null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC      function is not essential for viability. The ced-3 and ced-4 gene
CC      products may be used to develop agents for treating conditions
CC      characterised by cell deaths, such as myocardial infarction, stroke,
CC      degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC      infection, aging or hair loss.
SQ      Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;
```

Query Match 99.8%; Score 998; DB 9; Length 7653;
Best Local Similarity 99.9%; Pred.No. 0.00e+00;
Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;


```
FT      repeat_unit      /number= Intron_4
FT      4688..4719
FT      /tag= o
FT      /rpt_type= INVERTED
FT      /note= "Repeat 3"
FT      5221..5330
FT      repeat_unit
FT      /tag= p
FT      /rpt_type= INVERTED
FT      /note= "Repeat 3"
FT      5547..5760
FT      exon
FT      /tag= q
FT      /number= Exon_5
FT      5761..5814
FT      intron
FT      /tag= r
FT      /number= Intron_5
FT      5815..5942
FT      /tag= s
FT      /number= Exon_6
FT      5943..6297
FT      intron
FT      /tag= t
FT      /number= Intron_6
FT      6062..6138
FT      repeat_region
FT      /tag= u
FT      /rpt_type= INVERTED
FT      6298..6537
FT      exon
FT      /tag= v
FT      /number= Exon_7
FT      6434
FT      mutation
FT      /tag= w
FT      /note= "C>T, from allele n1129 and n1164"
FT      6538..7012
FT      intron
FT      /tag= x
FT      /number= Intron_7
FT      6567..6625
FT      repeat_unit
FT      /tag= y
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT      6905..6965
FT      repeat_unit
FT      /tag= z
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      7012..7075
FT      exon
FT      /tag= aa
FT      /number= Exon_8
FT      WO9325685-A.
PN      23-DEC-1993.
PD      14-JUN-1993; U05701.
PF      12-JUN-1992; US-897788.
PR      20-NOV-1992; US-979638.
PA      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI      Horvitz HR, Shaham S, Yuan J;
DR      WPI: 94-007542/01.
DR      P-PSDB: R53285.
PT      Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT      develop agents to increase or prevent cell death in organisms
PS      Claim 14; Fig 4; 127pp, English.
CC      The sequences given in Q64735-45 represent mutations of the C. elegans
CC      ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC      was most abundant in embryos, but was also detected in larvae and young
CC      adults, suggesting that ced-3 is expressed not only in cells undergoing
CC      cell death. The four largest introns as well as sequences 5' of the
CC      start codon contain repetitive elements, some of which have been
CC      characterised in non-coding regions of other C. elegans genes, such
CC      as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC      length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC      region that might be a transmembrane region. One region of Ced-3 is
CC      very rich in serine. It is thought that this region is involved in
CC      protein-protein interactions, similar to acid blobs in transcription
CC      factors. Of the mutations which occur within the ced-3 gene, eight of
CC      the mutations are missense mutations, two are nonsense mutations and
CC      two are putative splicing mutations. These mutations establish the
CC      null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC      function is not essential for viability. The ced-3 and ced-4 gene
```

```
CC      products may be used to develop agents for treating conditions
CC      characterised by cell deaths, such as myocardial infarction, stroke,
CC      degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC      infection, aging or hair loss.
SQ      Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.8%; Score 998; DB 9; Length 7653;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db      3601 tcatgtggttltcatatgatgacccaataatagcaatttttatatatattccctat 3660
QY      3601 TCATGTGTTTGTTCATATGATGACGAAAAATAGCAATTTTATATATTTCCCTAT 3660
Db      3661 tcatgttgcagaaaaatagtaaaaaagcgcatcatttttcgacatttttacatcga 3720
QY      3661 TCATGTGTGTCAGAAAAATAGTAAAAAGCGCATGCTTTTCGACATTTTTCATCGA 3720
Db      3721 acgacagctcacttcacatgctgaagaagagagcgaggaataataccacacatcttct 3780
QY      3721 ACGACAGCTCACTTCACATGCTGAAGACGAGAGCGGAGAAATACCAACATCTTCT 3780
Db      3781 gcgctctcgtcgttcacagatgtaaatggatctcggtcgatgtlaaaaaatgtcgata 3840
QY      3781 GCGTCTCTGCTTCAAGATGTGAATGGATCGGTCGATGTAAAAAATGTGGAATA 3840
Db      3841 atgtaaaaaatgcatgctgttttttacactttctgcacaaatgatataggggaaatgt 3900
QY      3841 ATGTAAAAAATGCATGCGTTTTTTTACACTTTTCGCACAAATGATAGGGGAAAAATGT 3900
Db      3901 attaaaaatacatttttttgtattttccaacatcacatgatattaaccattatlttcgtt 3960
QY      3901 ATTAAAAATACATTTTGTGATTTTTCACATTCACATGATTAACCCCATTTATTTTTCGTT 3960
Db      3961 gagcaacttlaaaagtagaagaatatagagcgaaacccaaatctcctaagatatacc 4020
QY      3961 GAGCAACTTAAAAAGTAGAAGATATTAGAGCGGAAAAACCAAAATTTCTTCAAGATATTACC 4020
Db      4021 ttattgataaattatagatgtlaataagcatatcttgaatgaagtcagcaaaaatagt 4080
QY      4021 TTTATTGATAATTATAGATGTTAATTAAGCATATCTTGAATGAAGTCAGCAAAAAATATGT 4080
Db      4081 gcgaaaaacacctgaaaaaatcaaaattctgcgaaaaatgaaaaaatgcatlaaataca 4140
QY      4081 GCGAAAAACACCTGAAAAAAATCAAAAATTCGCGAAAAATTGAAAAAATGCATTTAAATACA 4140
Db      4141 tttttgcatttttctacatcacatgtaatgtagaanaatlaaagggaatcaaaatttcta 4200
QY      4141 TTTTTCATTTTTCCTACATCACATGAATGTAGAAAAATTAAAGGGAATCAAAATTTCTA 4200
Db      4201 gaggatataattgaatgaanaacatttgcgaatlaaanaatgtgcgaacgcgcaaaaagaga 4260
QY      4201 GAGGATATATATTGAATGAAGAAACATTTGCGAAATTAATAATGTGCGAAACGTCAAAAAGAGA 4260
Db      4261 aatttgggtatcaaaatcgatctcctaacaacacacatttcagcatccgccaactctcat 4320
QY      4261 AATTGGGTATCAAAATCGATCCTTAAACCAACACATTTTACGATCCGCCAATCTTTCAT 4320
Db      4321 tcaccggatgctcttcttcggtacagttcaagtcgtaatcgctcattcagcaaaagctt 4380
QY      4321 TCACCGGATGCTCTCTCTCGGATACAGTTCAAGTGTATCGCTCAATCGCAAAAGCTT 4380
Db      4381 ctggaaccaactcaatacatattccatgaagagagatgaacttttcgattgcaccacaacca 4440
QY      4381 CTGGAACCAACTCAATATCATATTCCATGAAGAGATATGAACCTTTGTGATGCACCAACCA 4440
Db      4441 taagccgtgttttcgacgagaaacacatgtacagaactctcgagtcctcgtgaaatgt 4500
QY      4441 TAAGCCGTGTTTTCGACGAGAAAAACCATGTACAGAACTTCTCGAGTCCTGTGAATGT 4500
Db      4501 gccatcataataaatgaacactttgagcagatgccaacacggaatggtaccagggcg 4560
QY      4501 GCCTCATCATAAATATGAACACTTTGAGCAGATGCCAACACGGAATGGTACCAAGGCCG 4560
```


CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.8%; Score 998; DB 9; Length 7653;

Best Local Similarity 99.9%; Pred. No. 0.00e+00;

Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Db 3601 tcatgtgtttgttcataatgaacaaataagcaattttatatatttccctat 3660
    |||||||
Qy 3601 TCATGTGTTGTTTCATATGATGACGAAATAAGCAATTTTATATATTTCCCTAT 3660
    |||||||
Db 3661 tcatgtgttcagaaataatagtaaaagcgcatgttttcgacatttttaccatga 3720
    |||||||
Qy 3661 TCATGTGTTGACAGAAATAAGTAAAGCGCATGTCATTTTCGACATTTTTCATCGA 3720
    |||||||
Db 3721 acgacagctcacttcacatgctgaagagagagacgaggaataatcacacatcttct 3780
    |||||||
Qy 3721 ACGACAGCTCCTTCACATGCTGAAGACGACAGACGCGGAGAAATACACACATCTTCT 3780
    |||||||
Db 3781 gcgtctcgtcttcagcatgtgaatgggacgtcgtcgatgtaaaaaaatgtcgata 3840
    |||||||
Qy 3781 GCGTCTCTGCTCTTACGATGTGAATGGGATCTCGGTGATGTAATAAATGTGGAATA 3840
    |||||||
Db 3841 atgtaaaaaatgcatgctgttttttaccatttctgcacaaatgaataggggaaatgt 3900
    |||||||
Qy 3841 ATGTAAAAATGCGATGCGTTTTTTTACACTTTTCTGCACAAATGAATAGGGGAAATGT 3900
    |||||||
Db 3901 attaaatacatattttgtatttttcaacatcacatgattaaccattattttcgtt 3960
    |||||||
Qy 3901 ATTAATAATACATTTTGTATTTTCAACATCACATGATTAACCCATATTTTTCGTT 3960
    |||||||
Db 3961 gagcaacttaaaaagtagagatatattagagcgaaaccaaatttctcaagatatacc 4020
    |||||||
Qy 3961 GAGCAACTTAAAAAGTAGAGATATTAGAGGAAACCAAAATTTCTCAAGATATTACC 4020
    |||||||
Db 4021 ttattgataataatagatgttaataagcatacttgaaatgaatcagcaaaaatatgt 4080
    |||||||
Qy 4021 TTTATGATAATATAGATGTTAATAAGCATATCTTGAATGAAAGTCAGCAAAATATGT 4080
    |||||||
Db 4081 gcgaaacacactgaaaaaaatcaaaaatctgcgaaaaatgaaaaaatgcattaaataca 4140
    |||||||
Qy 4081 GCGAAACACCTGAAAAAATCAAAAATTCGCGAAAAATTGAAAAAATGCATTAAATACA 4140
    |||||||
Db 4141 ttttgcattttctacatcacatgaaatgtagaaaaatlaaaaggaatcaaaatttcta 4200
    |||||||
Qy 4141 TTTTGCATTTTCTACATCACATGAATGTAAGAAATTTAAAGGGAATCAAAATTTCTA 4200
    |||||||
Db 4201 gaggatataattgaatgaacatgtcgaaattaaaatgtgcgaacgtcaaaaaagaga 4260
    |||||||
Qy 4201 GAGGATATATGATGAAGAACATTCGGAATTAATAATGTGCGAAACGTCAAAAAAGAGA 4260
    |||||||
Db 4261 aatttgggtatcaaatcgatccttaaaacacacatttcagcatccgccaacttctcat 4320
    |||||||
Qy 4261 AATTGGGTATCAAAATCGATCCTAAAAACCAACACATTTTCAGCATCCGCCAATCTTCAT 4320
    |||||||
Db 4321 tcaccgatgtcttctctcgatagacagttcaagtctgtaatcgctcatcagcaaaagctt 4380
    |||||||
Qy 4321 TCACCGGATGCTCTCTCTCGGATACAGTCAAGTCGTAATCGTCAATTCAGCAAAAGCTT 4380
    |||||||
Db 4381 ctggaccaactcaatacatattccatgaagagatatgaacttgtcgatgcaccaacca 4440
    |||||||
```

```
Qy 4381 CTGACCACTCAATACATATTCATTCAGAGAGATATGAACCTTTGTCGATGCACCAACCA 4440
Db 4441 taagccgtgttttcgaagagaaacacatgttacagaactctcgcagtcctgtgaaatgt 4500
    |||||||
Qy 4441 TAAGCCGTGTTTTCGACGAGAAACCATGTACAGAACTTCTCGAGTCTCGTGAATGT 4500
    |||||||
Db 4501 gccatcataataatgaacacttgcagatgccaaacaggaatggtaccagagccg 4560
    |||||||
Qy 4501 GCCTCATCATTAATATGAACACTTTGAGCAGATGCCAACACGGAATGTTACCAAGGCCG 4560
    |||||||
Db 4561 acaagacaacttaccacattgttcagatgcagtgcta 4600
    |||||||
Qy 4561 ACAAGGACAATCTTACCAATTTGTCAGATGCATGGGCTA 4600
    |||||||

RESULT 9
ID Q64742 standard; DNA; 7653 BP.
AC Q64742;
DT 23-JUN-1994 (first entry)
DE ced-3 (C6434T) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
    /*tag= a
    /rpt_type= INVERTED
    /note= "Inverted w.r.t repeat at 1490-1614"
    1490..1614
    /*tag= b
    /rpt_type= INVERTED
    /note= "Inverted w.r.t. repeat at 1356-1472"
    2232..2366
    /*tag= c
    /number= Exon_1
    2367..2429
    /*tag= d
    /number= Intron_1
    2450..2575
    /*tag= e
    /number= Exon_2
    2576..2853
    /*tag= f
    /number= Intron_2
    2854..3107
    /*tag= g
    /number= Exon_3
    3108..4302
    /*tag= h
    /number= Intron_3
    3126..3243
    /*tag= i
    /rpt_type= INVERTED
    /note= "Inverted w.r.t. repeat at 3329-3396"
    3329..3396
    /*tag= j
    /rpt_type= INVERTED
    /note= "Inverted w.r.t. repeat at 3126-3243"
    3487..3759
    /*tag= k
    /rpt_type= INVERTED
    /note= "Inverted w.r.t. repeat at 3782-4070"
    3782..4070
    /*tag= l
    /rpt_type= INVERTED
    /note= "Inverted w.r.t. repeat 3487-3759"
    4303..4634
    /*tag= m
    /number= Exon_4
    4635..5546
    /*tag= n
```



```
|||||
QY 4261 AATTGGGTATCAAAATCGATCCTAAACCAACACATTTGAGCATCCGCCAACTCTTCAT 4320
Db 4321 tcaccgagatgctcttcctcgcgatacagttcaagtcgtaatcgctcattcagcaagctt 4380
QY 4321 TCACCGGATGCTCTCTCTCGGATACAGTTCAAGTCGTAATCGCTCATTTGAGCAAAAGCTT 4380
Db 4381 ctggaccaactcaatcacatattccatggaagagatatgaacttctcgatgcaccaca 4440
QY 4381 CTGACCAACTCAATACATATTCATGAGAGGATATGAACCTTGTTCGATGCACCAACCA 4440
Db 4441 taagccgtgttttcgacgagaaaaccatgtacagaacttctcgagtcctcgtggaatgt 4500
QY 4441 TAAGCCGTGTTTCGACGAGAAACCATGTACAGAAACTTCTCGAGTCTCTCGTGAATGT 4500
Db 4501 gccctacataataatgaacactttgagcagatgccacaacggaatggtaccaagccg 4560
QY 4501 GCCTCATCATAAATATGAACACTTTGAGCAGATGCCAACACGGAATGTTACCAAGGCCG 4560
Db 4561 acaagacaatctaccgaattgttcagatcgatggcta 4600
QY 4561 ACAAGGACAATCTTACCAATTTGTTGATGCATGGGCTA 4600

RESULT 8
ID 064745 standard; DNA; 7653 BP.
AC 064745;
DE 23-JUN-1994 (first entry)
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT FT /*tag= a
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t repeat at 1490-1614"
FT FT /*tag= b
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT FT /*tag= c
FT FT /number= Exon_1
FT FT /tag= d
FT FT /number= Intron_1
FT FT /number= 2450..2575
FT FT /*tag= e
FT FT /number= Exon_2
FT FT /tag= f
FT FT /number= Intron_2
FT FT /tag= g
FT FT /number= Exon_3
FT FT /tag= h
FT FT /number= Intron_3
FT FT /tag= i
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT FT /*tag= j
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT FT /*tag= k
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 3782-4070"

|||||
FT repeat_unit 3782..4070
FT FT /*tag= l
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 3487-3759"
FT FT /*tag= m
FT FT /number= Exon_4
FT FT /tag= n
FT FT /number= Intron_4
FT FT /tag= o
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT /*tag= p
FT FT /rpt_type= INVERTED
FT FT /note= "Repeat 3"
FT FT /*tag= q
FT FT /number= Exon_5
FT FT /tag= r
FT FT /number= Intron_5
FT FT /*tag= s
FT FT /number= Exon_6
FT FT /tag= t
FT FT /number= Intron_6
FT FT /*tag= u
FT FT /rpt_type= INVERTED
FT FT /tag= v
FT FT /number= Exon_7
FT FT /tag= w
FT FT /number= Intron_7
FT FT /*tag= x
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT FT /*tag= y
FT FT /rpt_type= INVERTED
FT FT /note= "Inverted w.r.t. repeat 6567-6625"
FT FT /*tag= z
FT FT /number= Exon_8
FT FT /tag= w
FT FT /note= "C>T, from allele n1163"

W09325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MAST ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shiham S, Yuan J;
DR WPI; 94-007542/01.
DR P-PSDB; R53288.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127pp; English.
CC The sequences given in Q64735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed as well as sequences 5' of the
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
```


FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 3329-3396"
FT	3329..3396
FT	/*tag= j
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 3126-3243"
FT	3487..3759
FT	/*tag= k
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 3782-4070"
FT	3782..4070
FT	/*tag= l
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat 3487-3759"
FT	4303..4634
FT	/*tag= m
FT	/number= Exon_4
FT	4635..5546
FT	/*tag= n
FT	/number= Intron_4
FT	4688..4719
FT	/*tag= o
FT	/rpt_type= INVERTED
FT	/note= "Repeat 3"
FT	5221..5330
FT	/*tag= p
FT	/rpt_type= INVERTED
FT	/note= "Repeat 3"
FT	5547..5760
FT	/*tag= q
FT	/number= Exon_5
FT	5761..5814
FT	/*tag= r
FT	/number= Intron_5
FT	5815..5942
FT	/*tag= s
FT	/number= Exon_6
FT	5943..6297
FT	/*tag= t
FT	/number= Intron_6
FT	5757
FT	/*tag= u
FT	/note= "C>T, from allele n717"
FT	6062..6138
FT	/*tag= v
FT	/rpt_type= INVERTED
FT	6298..6537
FT	/*tag= w
FT	/number= Exon_7
FT	6538..7012
FT	/*tag= x
FT	/number= Intron_7
FT	6567..6625
FT	/*tag= y
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 6905-6965"
FT	6905..6965
FT	/*tag= z
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat 6567-6625"
FT	7012..7075
FT	/*tag= aa
FT	/number= Exon_8
PN	WO9325685-A.
PU	23-DEC-1993.
PI	14-JUN-1993; U05701.
PR	12-JUN-1992; US-897788.
PA	20-NOV-1992; US-979638.
PI	(MASI) MASSACHUSETTS INST TECHNOLOGY.
PI	Horvitz HR, Shaham S, Yuan J;
DR	WPI; 94-007542/01.
PT	develop C elegans cell death genes ced-3 and ced-4 - used to
	isolated agents to increase or prevent cell death in organisms

PS	Claim 14; Fig 4; 127pp; English.
CC	The sequences given in Q64735-45 represent mutations of the C. elegans ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC	was most abundant in embryos, but was also detected in larvae and young
CC	adults, suggesting that ced-3 is expressed not only in cells undergoing
CC	cell death. The four largest introns as well as sequences 5' of the
CC	start codon contain repetitive elements, some of which have been
CC	characterised in non-coding regions of other C. elegans genes, such
CC	as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC	length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC	region that might be a transmembrane region. One region of Ced-3 is
CC	very rich in serine. It is thought that this region is involved in
CC	protein-protein interactions, similar to acid blobs in transcription
CC	factors. Of the mutations which occur within the ced-3 gene, eight of
CC	the mutations are missense mutations, two are nonsense mutations and
CC	two are putative splicing mutations. These mutations establish the
CC	null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC	function is not essential for viability. The ced-3 and ced-4 gene
CC	products may be used to develop agents for treating conditions
CC	characterised by cell deaths, such as myocardial infarction, stroke,
CC	degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC	infection, aging or hair loss.
SQ	Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;
	Query Match 99.8%; Score 998; DB 9; Length 7653;
	Best Local Similarity 99.9%; Pred. No. 0.00e+00;
	Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db	3601 tcatgtggttgttcatatgatgacccaaaatagcaattttatatatttccctat 3660
QY	3601 TCATGTGTTTGTTCATATGAATGACGAAAAATAGCAATTTTATATATTTTCCCTAT 3660
Db	3661 tcatgttgtgcagaaaaaatagtaaaaaagcgcatgcatatttcgacatttttaccatcga 3720
QY	3661 TCATGTGTGCAGAAAAATAGTAAAAAAGCGCATGCAATTTTCCACATTTTTACATCGA 3720
Db	3721 acgacagctcacttcacatgctgtaagacgagagacgcgagagaataccacacatcttct 3780
QY	3721 ACGACAGCTCACTTCACATGCTGAAGACGAGAGACGCGGAGAAATACACACATCTTCT 3780
Db	3781 gcgctctcgctctcagcatgtaaatggaatcgctcgatgtaaaaaaatgtcgaata 3840
QY	3781 GCGCTCTCGCTCTCAGCATGTGAATGGATCTCGTGCATGTAAAAAATGTGGAATA 3840
Db	3841 atgtaaaaaatgcatgcttttttaccatttctgcacaaatggaataggggaaaatgt 3900
QY	3841 ATGTAAAAAATGCATGCGTTTTTTTACACTTTCTGCACAAATGATAGGGGAAAAATGT 3900
Db	3901 attaaaatacattttttgtattttccaacatcacatgatttaaccattatttttcgtt 3960
QY	3901 ATTTAAATACATTTTGTGATTTTTCACATCACATGATTAACCCATTAATTTTCGTT 3960
Db	3961 gagcaacttaaaaaagttagaataatagaagcgaaaaaaccataatltcttcaagatatacc 4020
QY	3961 GAGCAACTTTAAAAAGTAGAGATAATTAGAGCGAAAAACCAAAATTTCTTCAAGATATTACC 4020
Db	4021 ttattgataaattatagatgttaataagcatatcttgaatgaaagtcagcaaaaatatgt 4080
QY	4021 TTATTTGATAATTATAGATGTTATTAAGCATATCTTGAATGAAAGTCAGCAAAAATATGT 4080
Db	4081 ggcgaacaccctgaaaaaaatcaaaaattctgcgaaaattgaaaaaatgcatlaaaatata 4140
QY	4081 GCGAACAACCTGAAAAAAATCAAAAATTCTGCGAAAATTGAAAAAATGCATTAAATACA 4140
Db	4141 tttttgcattttttctacatcacatgaaatgtagaanaattaaagggaatcaaaaattcta 4200
QY	4141 TTTTTCATTTTCTTCTACATCACATGAATGTAGAAAATTAAAAAGGGAATCAAAATTTCTA 4200
Db	4201 gaggataataattgaaatgaaacatgcygaaattaaaaatgtgcgaaacgltcaaaaaagagga 4260
QY	4201 GAGGATATAATTGAATGAACAATTCGGAATTTAAAAATGTGCGGAACGTCAAAAAAGAGGA 4260
Db	4261 aatttgggtatcaaaaatcgatccctaaaaaaccacacatttcagcatccgccaactcttcat 4320

PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI: 94-007542/01.
DR P-PSDB; R53287.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127pp; English.
CC The sequences given in Q64735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

Query Match 99.88; Score 998; DB 9; Length 7653;
Best Local Similarity 99.98; Pred. No. 0.00e+00;
Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 3601 tcatgtgtgttgcataatgaacaaatgcaattttatatatttccctat 3660
QY 3601 TCATGTGTTGTTGCATATGATGACGAAAAATAGCAATTTTATATATTTCCCTAT 3660
Db 3661 tcatgtgtgcagaaaaatagtaaaagcgcatgattttcgacattttacatcga 3720
QY 3661 TCATGTGTGCGAGAAAAATAGTAAAAAGCGCATGATTTTTCGACATTTTACATCGA 3720
Db 3721 acgacagctcacttcacatgctgaagacgagagcgagaaataaccacatcttct 3780
QY 3721 ACGACAGCTCACTTCACATGCTGAAGACGAGAGCGGAGAAATACACACATCTTCT 3780
Db 3781 gcgtctcgtcttcacgcatgtgaatggtatctcgatgttaaaaaatgtcgata 3840
QY 3781 GCGTCTCTGCTCTCAGCATGTGAATGGATCTCGTCGATGTAAAAAATGTGGAATA 3840
Db 3841 atgtaaaaaatgcatgctttttttacacttttctgcacaaatgaataggggaaaaatgt 3900
QY 3841 ATGTAAAAAATGCATGCGTTTTTCACACTTTCTGCACAAATGAATAGGGGAAAAATGT 3900
Db 3901 attaaaaatatt 3960
QY 3901 ATTAAAAATACATTTTGTGATTTTTCACATCAGATTAATTAATTAATTAATTAATTA 3960
Db 3961 gagcaacttaaaaagtagagataatlagagcgaaaaaatttctcaagatatacc 4020
QY 3961 GAGCAACTTAAAAAGTAGAGATATTAAGCGAAAAACCAAAATTTCTTCAAGATTTACC 4020
Db 4021 ttatgtataatataatgttaataagcatatcttgatgaagaaagtcagcaaaaatatgt 4080
QY 4021 TTTATGTATAATTATAGATGTTAATAAGCATATCTTGAATGAAGTCAGCAAAATATGT 4080
Db 4081 gcgaacacactgaaaaaaatcaaaaatctgcgaaaaatgaaaaaatgcatiaaaataca 4140
QY 4081 GCGAAACACCTGAAAAAATCAAAAATCTCTCGAAAAATGAAAAAATGATTAATAATACA 4140

Db 4141 ttttgcatttttctacatcacatgtaatgtagaanaattaaagggaatcaaaatttcta 4200
QY 4141 TTTTGCATTTTCTACATCACATGATGTAGAAAAATTAAAGGGAATCAAAATTTCTA 4200
Db 4201 gaggatataatgtaagaacattgcgaaattaaatgtgcgaaaacgtcaaaaaagagga 4260
QY 4201 GAGGATATAATTGAATGAACATTTGCCAAATTAATGTGCGAAACGTCAAAAAGAGGA 4260
Db 4261 aattggatcaaaatcgatccctaaacacacacatttcagcatccgccaacttcat 4320
QY 4261 AATTGGGATCAAAAATCGATCCTAAAAACCAACACATTTTCAGCATCCGCCAATCTTCAT 4320
Db 4321 tcaccgagatgctctcttcgagatcacagltcaagtcgtaatcgctcatcagcaaaagctt 4380
QY 4321 TCACCGGATGCTCTTCTCGGATACAGTTCAAGTCGTAATCGCTCATTCAGCAAAAGCTT 4380
Db 4381 ctggaaccaactcaatacatattccatgaagagatatgaacttgcgatgcaccaacca 4440
QY 4381 CTGGAACCAACTCAATACATATTCATGAAGAGGATATGAACCTTGTGATGACCAACCA 4440
Db 4441 taagccgtgttttcgagcgaaaaaccatgtacagaaaacttctcgagtcctcgtgaaatgt 4500
QY 4441 TAAGCCGTGTTTTCGACGAGAAACCATGTACAGAAACTTCTCGAGTCTCGTGAATGT 4500
Db 4501 gccctcatcataataatgaacactttgagcagatgccaacacggaatgttaccaagccg 4560
QY 4501 GCCTCATCATTAATAATATGACACTTTGAGCAGATGCCAACACGGAATGTATCAAGGCCG 4560
Db 4561 acaaggacaatcttaccattgttcagatgcatggcta 4600
QY 4561 ACAAGGACAACTTACCAATTTGTTTCAGATGATGGGCTA 4600

RESULT 7
ID Q64739 standard; DNA; 7653 BP.
AC Q64739;
DT 23-JUN-1994 (first entry)
DE ced-3 (G6297A) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT repeat_unit 1356..1472
FT FT /*tag= a
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat at 1490-1614"
FT repeat_unit 1490..1614
FT FT /*tag= b
FT /*rpt_type= INVERTED
FT /*note= "Inverted w.r.t. repeat at 1356-1472"
FT exon 2232..2366
FT FT /*tag= c
FT /*number= Exon_1
FT FT 2367..2429
FT FT /*tag= d
FT /*number= Intron_1
FT FT 2450..2575
FT FT /*tag= e
FT /*number= Exon_2
FT FT 2576..2853
FT FT /*tag= f
FT /*number= Intron_2
FT FT 2854..3107
FT FT /*tag= g
FT /*number= Exon_3
FT FT 3108..4302
FT FT /*tag= h
FT /*number= Intron_3
FT FT 3126..3243
FT FT /*tag= i

FT	/number= Exon_1
FT	2367..2429
FT	/*tag= d
FT	/number= Intron_1
FT	2450..2575
FT	/*tag= e
FT	/number= Exon_2
FT	2576..2853
FT	/*tag= f
FT	/number= Intron_2
FT	2854..3107
FT	/*tag= g
FT	/number= Exon_3
FT	3108..4302
FT	/*tag= h
FT	/number= Intron_3
FT	3126..3243
FT	/*tag= i
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 3329-3396"
FT	3329..3396
FT	/*tag= j
FT	/rpt_type= INVERTED
FT	/note= "inverted w.r.t. repeat at 3126-3243"
FT	3487..3759
FT	/*tag= k
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat at 3782-4070"
FT	3782..4070
FT	/*tag= l
FT	/rpt_type= INVERTED
FT	/note= "Inverted w.r.t. repeat 3487-3759"
FT	4303..4634
FT	/*tag= m
FT	/number= Exon_4
FT	4635..5546
FT	/*tag= n
FT	/number= Intron_4
FT	4688..4719
FT	/*tag= o
FT	/rpt_type= INVERTED
FT	/note= "Repeat 3"
FT	5221..5330
FT	/*tag= p
FT	/rpt_type= INVERTED
FT	/note= "Repeat 3"
FT	5547..5760
FT	/*tag= q
FT	/number= Exon_5
FT	5761..5814
FT	/*tag= r
FT	/number= Intron_5
FT	5815..5942
FT	/*tag= s
FT	/number= Exon_6
FT	5943..6297
FT	/*tag= t
FT	/number= Intron_6
FT	6062..6138
FT	/*tag= u
FT	/rpt_type= INVERTED
FT	6298..6537
FT	/*tag= v
FT	/number= Exon_7
FT	6372
FT	/*tag= w
FT	/note= "G>A, from allele n1286"
FT	6538..7012
FT	/*tag= x
FT	/number= Intron_7
FT	6567..6625
FT	/*tag= y
FT	/rpt_type= INVERTED

FT	repeat_unit	/note= "Inverted w.r.t. repeat at 6905-6965"
FT		6905..6965
FT		/*tag= 2
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat 6567-6625"
FT	exon	7012..7075
FT		/*tag= aa
FT		/number= Exon_8
PN	W09325685-A.	
PD	23-DEC-1993.	
PF	14-JUN-1993; U05701.	
PR	12-JUN-1992; US-897788.	
PR	20-NOV-1992; US-979638.	
PA	(MASI) MASSACHUSETTS INST TECHNOLOGY.	
PI	Horvitz HR, Shaham S, Yuan J;	
DR	WPI: 94-007542/01.	
DR	P-PSDB; R53284.	
PT	Isolated C elegans cell death genes ced-3 and ced-4 - used to	
PT	develop agents to increase or prevent cell death in organisms	
PS	Claim 14; Fig 4; 127pp; English.	
CC	The sequences given in 064735-45 represent mutations of the C. elegans	
CC	ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and	
CC	was most abundant in embryos, but was also detected in larvae and young	
CC	adults, suggesting that ced-3 is expressed not only in cells undergoing	
CC	cell death. The four largest introns as well as sequences 5' of the	
CC	start codon contain repetitive elements, some of which have been	
CC	characterised in non-coding regions of other C. elegans genes, such	
CC	as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in	
CC	length. Ced-3 is highly hydrophilic with no significant hydrophobic	
CC	region that might be a transmembrane region. One region of Ced-3 is	
CC	very rich in serine. It is thought that this region is involved in	
CC	protein-protein interactions, similar to acid blobs in transcription	
CC	factors. Of the mutations which occur within the ced-3 gene, eight of	
CC	the mutations are missense mutations, two are nonsense mutations and	
CC	two are putative splicing mutations. These mutations establish the	
CC	null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,	
CC	function is not essential for viability. The ced-3 and ced-4 gene	
CC	products may be used to develop agents for treating conditions	
CC	characterised by cell deaths, such as myocardial infarction, stroke,	
CC	degenerative disease, traumatic brain injury, hypoxia, pathogenic	
CC	infection, aging or hair loss.	
SQ	Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;	
Query Match	99.8%; Score 998; DB 9; Length 7653;	
Best Local Similarity	99.9%; Pred. No. 0.00e+00;	
Matches 999; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
Db	3601 tcatgtgtttgttcataatgaatgacccaaaatagcaatttttatataatttccctat	3660
QY	3601 TCATGTGTTTGTTCATATGAATGACGCAAAATAGCAATTTTATATTTTCCCTAT	3660
Db	3661 tcatgttgcagaaaaaatagtaaaaaagcgcatgcatatttcgacatttttaccatga	3720
QY	3661 TCATGTGTGCAGAAAAATAGTAAAAAAGCGCATGCATTTTTCGACATTTTTCACATCGA	3720
Db	3721 acgacagctcaattcacatgctgaagacgagagagacgcgagagaataaccacacatcttct	3780
QY	3721 ACGACAGCTCACTTCACATGCTGAAGACGAGAGACGGGAGAAATACACACATCTTCT	3780
Db	3781 gcgtctcgtcttcagcatgtgaatgagatctcggtcgatgttaaaaaaatgtcgata	3840
QY	3781 GCGTCTCTGCTCTTCAGCATGTGAATGGATCTCGGTGATGTAAAAAATGTGGAATA	3840
Db	3841 atgtaaaaatcatalgcgttttttaccattttctgcacaaatgaatagggggaaaatgt	3900
QY	3841 ATGTAAAAATGCATGCGTTTTTTTACACTTTTCTGCACAAATGAATAGGGGGAATAATGT	3900
Db	3901 attaaatatcatttttgtatttttcaacatcacatgattaaaccocatatttttcgtt	3960
QY	3901 ATTAATAATACATTTTGTATTTTTCAACATCACATGATTAACCCCATATTTTTTCGTT	3960
Db	3961 gagcaacttaaaaagtagagatatattagagcgaaaaaccaaatttcttcaagatatacc	4020


```

FT      exon      6298..6537
FT      /tag= w
FT      /number= Exon_7
FT      intron    6538..7012
FT      /tag= x
FT      /number= Intron_7
FT      repeat_unit 6567..6625
FT      /tag= y
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT      repeat_unit 6905..6965
FT      /tag= z
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      exon      7012..7075
FT      /tag= aa
FT      /number= Exon_8
PN      WO9325685-A.
PD      23-DEC-1993.
PF      14-JUN-1993; U05701.
PR      12-JUN-1992; US-897788.
PR      20-NOV-1992; US-979638.
PA      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI      Horvitz HR, Shaham S, Yuan J;
DR      WPI; 94-007542/01.
DR      P-PSDB; R53282.
PT      Isolated C elegans cell death genes ced-3 and ced-4 - used to
PS      develop agents to increase or prevent cell death in organisms
PS      Claim 14; Fig 4; 127pp; English.
CC      The sequences given in Q64735-45 represent mutations of the C. elegans
CC      ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC      was most abundant in embryos, but was also detected in larvae and young
CC      adults, suggesting that ced-3 is expressed not only in cells undergoing
CC      cell death. The four largest introns as well as sequences 5' of the
CC      start codon contain repetitive elements, some of which have been
CC      characterised in non-coding regions of other C. elegans genes, such
CC      as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC      length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC      region that might be a transmembrane region. One region of Ced-3 is
CC      very rich in serine. It is thought that this region is involved in
CC      protein-protein interactions, similar to acid blobs in transcription
CC      factors. Of the mutations which occur within the ced-3 gene, eight of
CC      the mutations are missense mutations, two are nonsense mutations and
CC      two are putative splicing mutations. These mutations establish the
CC      null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC      function is not essential for viability. The ced-3 and ced-4 gene
CC      products may be used to develop agents for treating conditions
CC      characterised by cell deaths, such as myocardial infarction, stroke,
CC      degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC      infection, aging or hair loss.
SQ      Sequence 7653 BP; 2429 A; 1452 C; 1272 G; 2500 T;

Query Match 99.8%; Score 998; DB 9; Length 7653;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      3841 ATGTAAATAATGCATGCGTTTCTTACACTTTCTGCACAAATGATAGGGGAAATGT 3900
Db      3901 attaaatacatatttttctgatatlttccaacatcacatgatlttaaccccatatttttcgtt 3960
QY      3901 ATTAAATACATTTTGTGATTTTTCACATCACATGATTAACCCCATTAFTTTTTCGTT 3960
Db      3961 gagcaacttaaaaagttagaataattagagcgaaacccaaaattcttcaagatatacc 4020
QY      3961 GAGCAACTTAAAGTAGAGATATTAGAGCGAAACCAAAATTTCTCAAGATATTACC 4020
Db      4021 ttattgataattatagatgttaataagcatatcttgaatgaagtgcagcaaaatatgt 4080
QY      4021 TTTATTGATAATTATAGATGTTAATAAGCATATCTTGAATGAAGTCAAGCAAAATATGT 4080
Db      4081 gcgaacacacctgaaaaaatcaaaaattctgcgaaattgaaaaaatgcattaaatata 4140
QY      4081 GCGAAACACCTGAAATAATCAAAAATTCGCGAAATAATGAAAAAATGCATTAAATACA 4140
Db      4141 ttttgcatttttctacatcacatgtaatgtagaataattaaagggaataatctcta 4200
QY      4141 TTTTGCATTTTCTACATCACATGATGTGCAAAATTTAAAGCGAAATCAAAATTTCTA 4200
Db      4201 gaggatataattgaatgaacacattgcgaaattaaatgtgcgaaacgtcaaaaaagaga 4260
QY      4201 GAGGATATAATTGATGAACATTCGCGAAATTTAAATGTGCGAAACGTCAAAAAGAGGA 4260
Db      4261 aattggtatcaaaaatcgatcctaataacacacatttcagcatccgccaactctcat 4320
QY      4261 AATTGGGTATCAAAATCGATCCTTAAACCAACACATTTCAGCATCCGCCAACTCTTCA 4320
Db      4321 tcaccggtctctctctcgcgatacagttcaagtcgtaatcgctcaatcagcaaaagct 4380
QY      4321 TCACCGGATGCTCTCTCTCGGATACAGTCAAGTCGTAATCGCTCAATTCAGCAAAAGCT 4380
Db      4381 ctgagcaactcaatcacatatctcatgaaagagatatgaactttgtcgatgcaccaacca 4440
QY      4381 CTGAGCAACTCAATACATATTCATGAAGAGATATGAACCTTGTGATGCACCAACCA 4440
Db      4441 taagccgtgttttcgcagcgaagaaacacatgtacagaactctcgaagtcctgtgaatgt 4500
QY      4441 TAAGCCGTGTTTTCGACGAGAAACCATGTACAGAAACTTCTCGAGTCTCTGGAATGT 4500
Db      4501 gccatcatataataatgaacacattgaagcagatgccaacgcgaatgttaccgaagccg 4560
QY      4501 GCCTCATATATAATAATGAACACCTTGAGCAGATGCCAACGCAATGTATCAAGGCCG 4560
Db      4561 acaagacaacttaccgaattgttcagatgcagtgcta 4600
QY      4561 ACAAGACAATCTTACCAATTTGTCAGATGCATGGGCTA 4600

RESULT 5
ID      Q64741; standard; DNA; 7653 BP.
AC      Q64741;
DE      23-JUN-1994 (first entry)
KW      C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW      embryogenesis; cell death; hydrophilic; transmembrane; region;
KW      hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW      protein synthesis; null phenotype; calcium-binding domain; ss.
OS      Caenorhabditis elegans.
FH      Key
FT      Location/Qualifiers
FT      repeat_unit 1356..1472
FT      /tag= a
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t repeat at 1490-1614"
FT      repeat_unit 1490..1614
FT      /tag= b
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 1356-1472"
FT      exon      2232..2366
FT      /tag= c

```


23-JUN-1994 (first entry)
DE ced-3 (G2487A) gene.
KW C. elegans; ced-4; ced-3; mutant; transcriptional regulation;
KW embryogenesis; cell death; hydrophilic; transmembrane; region;
KW hydrophobic; mutation; amino acid; substitution; RNA splicing;
KW protein synthesis; null phenotype; calcium-binding domain; ss.
OS Caenorhabditis elegans.
FH key location/Qualifiers
FT repeat_unit 1356..1472
FT /tag= a
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t repeat at 1490-1614"
FT 1490..1614
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 1356-1472"
FT 2232..2366
FT /tag= c
FT /number= Exon_1
FT 2367..2429
FT /tag= d
FT /number= Intron_1
FT 2450..2575
FT /tag= e
FT /number= Exon_2
FT 2487
FT /tag= f
FT /note= "G>A, from allele n718"
FT 2576..2853
FT /tag= g
FT /number= Intron_2
FT 2854..3107
FT /tag= h
FT /number= Exon_3
FT 3108..4302
FT /tag= i
FT /number= Intron_3
FT 3126..3243
FT /tag= j
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3329-3396"
FT 3329..3396
FT /tag= k
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3126-3243"
FT 3487..3759
FT /tag= l
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 3782-4070"
FT 3782..4070
FT /tag= m
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 3487-3759"
FT 4303..4634
FT /tag= n
FT /number= Exon_4
FT 4635..5546
FT /tag= o
FT /number= Intron_4
FT 4688..4719
FT /tag= p
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5221..5330
FT /tag= q
FT /rpt_type= INVERTED
FT /note= "Repeat 3"
FT 5547..5760
FT /tag= r
FT /number= Exon_5
FT 5761..5814
FT /tag= s
FT /number= Intron_5

FT exon 5815..5942
FT /tag= t
FT /number= Exon_6
FT 5943..6297
FT /tag= u
FT /number= Intron_6
FT 6062..6138
FT /tag= v
FT /rpt_type= INVERTED
FT 6298..6537
FT /tag= w
FT /number= Exon_7
FT 6538..7012
FT /tag= x
FT /number= Intron_7
FT 6567..6625
FT /tag= y
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat at 6905-6965"
FT 6905..6965
FT /tag= z
FT /rpt_type= INVERTED
FT /note= "Inverted w.r.t. repeat 6567-6625"
FT 7012..7075
FT /tag= aa
FT /number= Exon_8

W09325685-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05701.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-979638.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007542/01.
DR P-PSDB; R53280.
PT Isolated C elegans cell death genes ced-3 and ced-4 - used to
PT develop agents to increase or prevent cell death in organisms
PS Claim 14; Fig 4; 127pp; English.
CC The sequences given in Q64735-45 represent mutations of the C. elegans
CC ced-3 gene. A 2.8 kb mRNA was identified as the ced-3 transcript and
CC was most abundant in embryos, but was also detected in larvae and young
CC adults, suggesting that ced-3 is expressed not only in cells undergoing
CC cell death. The four largest introns as well as sequences 5' of the
CC start codon contain repetitive elements, some of which have been
CC characterised in non-coding regions of other C. elegans genes, such
CC as fem-1, lin-12 and myoD. The Ced-3 protein is 503 amino acids in
CC length. Ced-3 is highly hydrophilic with no significant hydrophobic
CC region that might be a transmembrane region. One region of Ced-3 is
CC very rich in serine. It is thought that this region is involved in
CC protein-protein interactions, similar to acid blobs in transcription
CC factors. Of the mutations which occur within the ced-3 gene, eight of
CC the mutations are missense mutations, two are nonsense mutations and
CC two are putative splicing mutations. These mutations establish the
CC null phenotype of the ced-3 gene, confirming that ced-3, like ced-4,
CC function is not essential for viability. The ced-3 and ced-4 gene
CC products may be used to develop agents for treating conditions
CC characterised by cell deaths, such as myocardial infarction, stroke,
CC degenerative disease, traumatic brain injury, hypoxia, pathogenic
CC infection, aging or hair loss.
SQ Sequence 7653 BP; 2430 A; 1453 C; 1271 G; 2499 T;

Query Match 99.8%; Score 998; DB 9; Length 7653;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 999; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 3601 tcatgtgttgcataatgaacaaataatgaacattttatataatccctat 3660
|||||
Qy 3601 TCATGTGTTGTCATATGAATGACGAAATAATGCAATTTTATATATTTCCCTAT 3660
|||||
Db 3661 tcatgttgcagaaataatagcagcagcatgtattttgcacatttttacatcga 3720
|||||
Qy 3661 TCATGTGTCAGAAATAATAGTAAATAAGCGCATGCAATTTTTCGACATTTTACATCGA 3720


```
FT intron 5943..6297
FT /tag= x
FT /label= Intron-6
FT repeat_unit 6062..6138
FT /tag= y
FT /label= Repeat-4
FT exon 6298..6537
FT /tag= z
FT allele 6322
FT /tag= aa
FT /label= T(n1949)
FT /note= "creates premature stop codon at 412"
FT allele 6372
FT /tag= ab
FT /label= A(n1286)
FT allele 6434
FT /tag= ac
FT /label= T(n1129,n1164)
FT /note= "causes A449V mutation"
FT allele 6485
FT /tag= ad
FT /label= T(n2430)
FT /note= "causes A466V mutation"
FT allele 6535
FT /tag= ae
FT /label= A(n2426)
FT /note= "causes E483K mutation"
FT intron 6538..7012
FT /tag= af
FT /label= Intron-7
FT repeat_unit 6567..6625
FT /tag= ag
FT /label= Repeat-5
FT /tag= ah
FT /label= Repeat-5
FT /tag= ai
FT /label= Repeat-5
FT /tag= aj
FT /label= T(n1163)
FT /note= "causes S486F mutation"
FT allele 7020
FT /tag= aj
FT /label= T(n1163)
FT /note= "causes S486F mutation"
PN WO9625946-A1.
PD 29-AUG-1996.
PF 23-FEB-1996; U02473.
PR 24-FEB-1995; US-394189.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR P-PSDB: R98754.
DR P-PSDB: R98754.
PT Ccd-3 and human interleukin 1-beta convertase genes and proteins
PT useful to treat inflammation and diseases characterised by cell
PT death
PS Claim 18; Fig 3; 139pp; English.
CC The Caenorhabditis elegans ced-3 gene (T38196) was cloned by
CC mapping RFLPs and chromosome walking, and genomic DNA cloned
CC in plasmid pJ107 was sequenced. EMS-induced alleles were also
CC sequenced. The gene codes for a cell death protein (R98754) that
CC is structurally similar to human interleukin-1 beta converting
CC enzyme (ICE) (R98755), suggesting that Ced-3 protein may be a
CC cysteine protease like ICE and that ICE may be a human equivalent
CC of the nematode cell death gene. The ced-3 gene can be used as a
CC probe or in the prodn. of Ced-3 protein and novel drugs for
CC enhancing or inhibiting the activity of ICE, ced-3 and related
CC genes for the treatment of inflammatory diseases and/or diseases
CC caused by cell death. Novel inhibitors of ced-3 activity include
CC portions of the ced-3 gene and its product.
SQ Sequence 7653 BP; 2429 A; 1456 C; 1270 G; 2498 T;
```

Query Match 100.0%; Score 1000; DB 24; Length 7653;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db 3601 tcatgtgtttgttcataatgaatgaacgaataatagcaattttatatatttccctat 3660
QY 3601 TCATGTGTTGTTCTCATATGAATGACGAAAAATAGCAATTTTATATATTTCCCTAT 3660
Db 3661 tcatgtgtgtcagaataatagtaaaaaagcgatgcattttcgacattttttacatcga 3720
QY 3661 TCATGTGTGACGAAAAATAGTAATAAAGCGCATGCAATTTTCGACATTTTACATCGA 3720
Db 3721 acgacagctcacttcacatgctgaagacgagagacgaggaataataccacacatcttct 3780
QY 3721 ACGACAGCTCACTTCACATGCTGAAGACGAGAGACGCGGAGAAATACACACATCTTCT 3780
Db 3781 gcgtctcgtcttcagcatgtgaatggatcgcgtcgatgttaaaaaatgtcgata 3840
QY 3781 GCGTCTCTGCTCTTCAGCATGTGAATGGGATCTCGGTGATGTAAAAAATGTGATA 3840
Db 3841 atgtaaaaatgcatgcgtttttttacactttctgcacaaatgaataggggaaatgt 3900
QY 3841 ATGTAAAAATGCATGCGTTTTTTTACACTTTCTGCACAAATGAATAGGGGAAATGT 3900
Db 3901 attaaatacatattttgtattttcaacatcacatgatatacccatatttttgcgt 3960
QY 3901 ATTAATAATACATTTTGTATTTTCAACATCACATGATTAACCCCATTTATTTTGTG 3960
Db 3961 gagcaacttaaaaaagtagagaatatagagcgcaaaaccaaatttcttcaagatatacc 4020
QY 3961 GAGCAACTTAATAAGTAGAGATATTAGAGCGAAAAACCAAAATTTCTCAAGATATTACC 4020
Db 4021 ttattgataattatagatgttaataagcatatcttgatgaagaagtcagcaaaatagt 4080
QY 4021 TTTATTGATAATTATAGATGTTAATAGCATATCTTGAATGAAGTCAGCAAAATATTGT 4080
Db 4081 gcgaacacccgtaaaaaaatcaaaattctgcgaaaatgtaaaaaatgcatlaaataca 4140
QY 4081 GCGAAACACCTGAAAAAATCAAAATTTCTGCGAAAAATTGAAAAATGCATTAAAAATACA 4140
Db 4141 tttttgcatttttctacatcacatgaatgtagaataatgaagggaatcaaatctta 4200
QY 4141 TTTTGCATTTTCTACATCACATGATGTAGAAAAATTAAAGGGAATCAAAATTTCTA 4200
Db 4201 gaggatataattgaatgaacattgcgaataatgaatgtgcgaacacgtaaaaaagaga 4260
QY 4201 GAGGATATAATTGAATGAACATTTGCGAAATTTAAATGTGCGCAAAACGTCAAAAAAGAGA 4260
Db 4261 aattgggtatcaaaaatcgatccttaaaaccaaacaatttcaagcatccgccaacttcat 4320
QY 4261 AATTGGGTATCAAAAATCGATCCTTAACCAACACATTTTCAGCATCCGCCAACTCTTCA 4320
Db 4321 tcaccgtagctcttctcctcgatataagttcaagtcgtaactcgctcatcagcaaaagct 4380
QY 4321 TCACCGGATGCTCTTCTCTCGATACAGTTCAAGTCGTAATCGCTCATTCAGCAAAAGCT 4380
Db 4381 ctggaaccaactcaatcatatccatgaagagatgatgaacttgcgatgcaccaacca 4440
QY 4381 CTGGAACCACTCAATACATATTCATGAAGAGATATGAATTTGTGATGACCAACCA 4440
Db 4441 taagccgtgttttcgacgagaaacacatgttacagaacttctcagagtcctcgtgaatgt 4500
QY 4441 TAAGCCGTGTTTTCGACGAGAAACCATGTACAGAACTTCTCGAGTCTCTGGAATGT 4500
Db 4501 gcctcatcataataatgaacactttgagcagatgccaaacggaatggtaccagaagcgg 4560
QY 4501 GCCTCATCATATAATATGAACACTTTGAGCAGATGCCAACGGAATGTTACCAAGGCGG 4560
Db 4561 acaagacaatcttacaattgttcagatgcattgctga 4600
QY 4561 ACAAGGACAACTTACCAATTGTTTCAGATGCGGCTA 4600
```

RESULT 3
ID Q64736 standard; DNA; 7653 BP.
AC Q64736;

|||||
QY 3601 TCATGTGTTGTTTCATATGATGACGAAAAATAGCAATTTTATATATTTTCCCTAT 3660
Db 3661 tcatgttgcagaaaaaatagtaaaaaagcgcatgcatlcttcgacatlltttacatcga 3720
QY 3661 TCATGTTGTGAGAAAAATAGTAAAAAAGCGCATGATTTTTCGACATTTTTCATCGA 3720
Db 3721 acgacagctcacttcacatgctgaagacgagagacgagagaataacacacatcttct 3780
QY 3721 ACGACAGCTCAGTTCACATGCTGAAGACGAGAGACGCGAGAAATACACACATCTTCT 3780
Db 3781 gcgtctcgtctcagcatgtgaatgggatctcgatgtaaaaaaatgtcgaata 3840
QY 3781 GCGTCTCTCGTCTCAGCATGTGAATGGGATCTCGGTCGATGTAAAAAATGTGAATA 3840
Db 3841 atgtaaaaaatgcatgctlltttacactllctgcacaaatgatataggggaaaaatgt 3900
QY 3841 ATGTAAAAAATGCATGCGTTTTTCACACTTTCTGCACAAATGAATAGGGGAAAAATGT 3900
Db 3901 attaaatacatlltttgtatllctcaacatcacatgattaacccattatlltctgt 3960
QY 3901 ATTTAAATACATTTTGTATTTTTCACATCACATGATTAAACCCATTATTTTTCGTT 3960
Db 3961 gagcaacttaaaagtagaataatagagcgaaaccaaattlctcaagatatacc 4020
QY 3961 GAGCAACTTAAAAAGTAGAGAATATTAGAGCGAAAAACCAAAATTTCTTCAAGATATTACC 4020
Db 4021 ttattgataatatagatgttaataagcatatcttgaatgaagtcagcaaaaatatgt 4080
QY 4021 TTTATTGATAATATATAGATGTTAATAAGCATATCTTGAATGAAGTCAGCAAAAAATATGT 4080
Db 4081 gcgaacacacctgaaaaaatcaaaaattctgcgaaaaattgaaaaaatgcatlaaataca 4140
QY 4081 GCGAAACACCTGAAAAAATCAAAAATCTGCGAAAAATTGAAAAAATGCATTTAAATACA 4140
Db 4141 ttttgcattlctctacatcacatgaaatgtagaaaaattaaagggaatcaaaaattcta 4200
QY 4141 TTTTGCATTTTCTACATCACATGAATGTAGAAAAATTAAAGGGAATCAAAAATTTCTA 4200
Db 4201 gaggatataatgaaatgaacatctgcgaatlaaaatgtgcgaacgtaaaaaagaga 4260
QY 4201 GAGGATATAATTGAAATGAACAATGCGAAATTAAATGTGCGAAACGTCAAAAAAGAGA 4260
Db 4261 aattgggtatcaaaaatcgatcctaacaacacacatttcagcatccgccaactctcat 4320
QY 4261 AATTGGGTATCAAAAATCGATCCTAAAAACCAACACATTTTCAGCATCCGCCAACTCTTCA 4320
Db 4321 tcaccgcatgctctctctcgcgatatcagttcaagtctgaatcgctcatcagcaaaagct 4380
QY 4321 TCACCGGATGCTCTCTCTCGGATACAGTTCGAAGTCGTAATCGCTCATTCAGCAAAAGCTT 4380
Db 4381 ctggaaccaactcaatatcatlctcatgaaagagatatgaactlctgcgatgcaccaacca 4440
QY 4381 CTGGACCAACTCAATACATATTCCATGAAGAGGATATGAACCTTTGTGATGCACCAACCA 4440
Db 4441 taagccgtgttttcgacgagaaaaacacatgtacagaaactlctcgagtcctcgtgaaatgt 4500
QY 4441 TAAGCCGTGTTTTCGACGAGAAAAACCATGTACAGAAACTTCTCGAGTCTCTGGAATGT 4500
Db 4501 gcctcatcataaataatgaacactttgagcagatgccaacacggaatggtaccagcg 4560
QY 4501 GCCTCATCATAAATATGAACACTTTGAGCAGATGCCAACACGGAATGTGTACCAAGGCCG 4560
Db 4561 acaagacaatcttaaccaatltgttcagatgcattggcta 4600
QY 4561 ACAAGGACAACTTACCAATTGTTCAGATGCATGGGCTA 4600

RESULT 2
ID T38196 standard; DNA; 7653 BP.
AC T38196;
DT 17-DEC-1996 (first entry)
DE Nematode Ced-3 gene.

KW Ced-3; interleukin-1 beta converting enzyme; ICE; protease;
KW cell death; apoptosis; neural degeneration; inflammation;
KW antiinflammatory; ds.
OS Caenorhabditis elegans.
FH Key location/Qualifiers
FT repeat_unit 1356..1472
FT /tag= a
FT /label= Repeat-1
FT /tag= b
FT /label= Repeat-2
FT /tag= c
FT /codon_start= 2232..2234
FT 2310
FT /tag= d
FT /label= T(n1040)
FT /note= "causes L27F mutation"
FT 2367..2429
FT /tag= e
FT /label= Intron-1
FT 2430..2575
FT /tag= f
FT 2487
FT /tag= g
FT /label= A(n718)
FT /note= "causes G65R mutation"
FT 2576..2853
FT /tag= h
FT /label= Intron-2
FT 2854..3107
FT /tag= i
FT 3108..4302
FT /tag= j
FT /label= Intron-3
FT 3126..3243
FT /tag= k
FT /label= Repeat-1
FT 3329..3396
FT /tag= l
FT /label= Repeat-1
FT 3487..3759
FT /tag= m
FT /label= Repeat-2
FT /tag= n
FT /label= Repeat-2
FT 4303..4633
FT /tag= o
FT 4634..5546
FT /tag= p
FT /label= Intron-4
FT 4688..4719
FT /tag= q
FT /label= Repeat-3
FT 5221..5330
FT /tag= r
FT /label= Repeat-3
FT 5546..5760
FT /tag= s
FT 5757
FT /tag= t
FT /label= A(n2433)
FT /note= "causes G360S mutation"
FT 5761..5814
FT /tag= u
FT /label= Intron-5
FT 5815..5942
FT /tag= v
FT 5940
FT /tag= w
FT /label= T(n1165)
FT /note= "creates premature stop codon at 403"

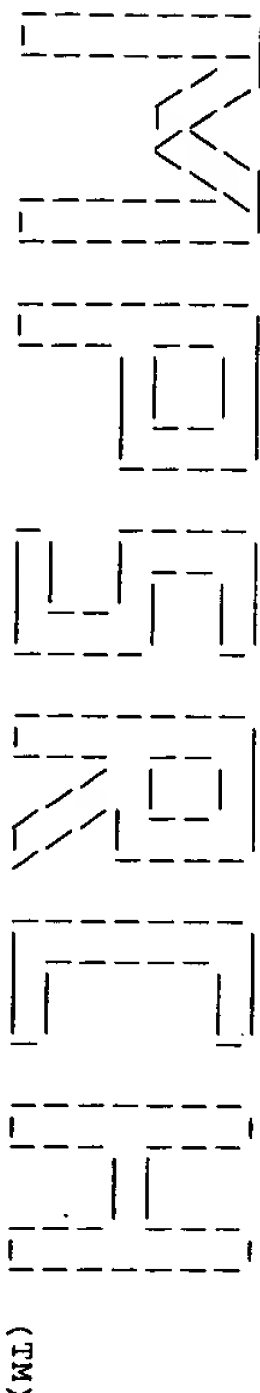

```
FT      /tag= g
FT      /note= "G>A, fromm allele n718"
FT      /tag= h
FT      /number= Intron_2
FT      /tag= i
FT      /number= Exon_3
FT      /tag= j
FT      /number= Intron_3
FT      /tag= k
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3329-3396"
FT      /tag= l
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3126-3243"
FT      /tag= m
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 3782-4070"
FT      /tag= n
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 3487-3759"
FT      /tag= o
FT      /number= Exon_4
FT      /tag= p
FT      /number= Intron_4
FT      /tag= q
FT      /rpt_type= INVERTED
FT      /note= "Repeat 3"
FT      /tag= r
FT      /rpt_type= INVERTED
FT      /note= "Repeat 3"
FT      /tag= s
FT      /number= Exon_5
FT      /tag= t
FT      /note= "G>A, from allele n2433"
FT      /tag= u
FT      /number= Intron_5
FT      /tag= v
FT      /number= Exon_6
FT      /tag= w
FT      /note= "C>T, from allele n1165"
FT      /tag= x
FT      /number= Intron_6
FT      /tag= y
FT      /rpt_type= INVERTED
FT      /tag= z
FT      /note= "G>A, fromm allele n717"
FT      /tag= aa
FT      /number= Exon_7
FT      /tag= ab
FT      /note= "C>T, from allele n1949"
FT      /tag= ac
```

```
FT      /note= "G>A, from allele n1286"
FT      /tag= ad
FT      /note= "C>T, fromm alleles n1129 and n1164"
FT      /tag= ae
FT      /note= "C>T, from allele n2430"
FT      /tag= af
FT      /note= "G>A, from allele n2426"
FT      /tag= ag
FT      /number= Intron_7
FT      /tag= ah
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat at 6905-6965"
FT      /tag= ai
FT      /rpt_type= INVERTED
FT      /note= "Inverted w.r.t. repeat 6567-6625"
FT      /tag= aj
FT      /number= Exon_8
FT      /tag= ak
FT      /note= "C>T, from allele n1163"
FT      /tag= al
FT      /note= "Claim 2, inhibitor fragment"
FT      /tag= am
FT      /note= "Claim 2, inhibiotr fragment"
```

```
PN      WO9325694-A.
PD      23-DEC-1993.
PF      14-JUN-1993; U05705.
PR      12-JUN-1992; US-8977788.
PR      20-NOV-1992; US-984182.
PA      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI      Horvitz HR, Shaham S, Yuan J;
DR      WPI: 94-007551/01.
DR      P-P5DB; R45262.
PT      Agents which affect activity of cell death genes - used to
PT      develop drugs for treating conditions characterised by cell death
PT      or proliferation
PS      Disclosure; Fig 3; 132pp; English.
CC      This sequence encodes the C.elegans cell death gene, ced-3. Fragments
CC      of the amino terminal of the protein encoded by this sequence act as
CC      inhibitors of ced-3. This gene has considerable similarity to human
CC      interleukin-1beta convertase (ICE), which converts pro-interleukin-
CC      1beta to the active cytokine and is involved in inflammatory response
CC      in humans. The similarity between the two sequences suggests that
CC      inhibitors of ced-3 may also act as inhibitors of ICE. Human ICE and
CC      nematode Ced-3 proteins have an overall amino acid similarity of 28%.
CC      The ced-3 inhibitors may be used for identifying agents which affect
CC      the activity of a gene belonging to the to the ced-3/ICE family of
CC      genes and for diagnosis of diseases characterised by cell death. They
CC      can also be used to develop drugs for treating conditions characterised
CC      by cell deaths such as myocardial infarction, stroke, degenerative
CC      disease, traumatic brain injury, hypoxia, pathogenic infection, or
CC      hair loss, or drugs for reducing the proliferative capacity or size
CC      of a population of cells such as cancerous cells, cells which produce
CC      autoreactive antibodies, infected cells, hair follicle cells or cells
CC      which are critical to the life of a parasite, pest or recombinant
CC      organism. They may also be used in the diagnosis of inflammatory
CC      disease.
SQ      Sequence      7653 BP;      2429 A;      1452 C;      1272 G;      2500 T;
```

```
Query Match      100.0%; Score 1000; DB 9; Length 7653;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Db 3601 tcatgtgttgcataatgacgaataatgcaattttatatatttccctat 3660



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sat Aug 7 19:20:42 1999; Maspar time 234.95 Seconds

Tabular output not generated. 911.985 Million cell updates/sec

Title: >US-08-287-669-18
Description: (3601-4600) from US08287669.seq (5 of 10)
Perfect Score: 1000
N.A. Sequence: 3601 TCATGTGTTTGTTCATATG.....TTGTCAGATGATGGGCTA 4600
Comp: AGTACACCAACAAGTATAC.....AACAGTCTACGTACCCGAT

Scoring table: TABLE default
Gap 6

Mmatch STD: Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

n-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60

Statistics: Mean 8.973; Variance 6.203; scale 1.447

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1000	100.0	7653	9	Q54401	ced-3 gene.	0.00e+00
2	1000	100.0	7653	24	T38196	Nematode Ced-3 gene.	0.00e+00
3	998	99.8	7653	9	Q64736	ced-3 (G2487A) gene.	0.00e+00
4	998	99.8	7653	9	Q64738	ced-3 (C5940T) gene.	0.00e+00
5	998	99.8	7653	9	Q64741	ced-3 (G6372A) gene.	0.00e+00
6	998	99.8	7653	9	Q64744	ced-3 (G6536A) gene.	0.00e+00
7	998	99.8	7653	9	Q64739	ced-3 (G6297A) gene.	0.00e+00
8	998	99.8	7653	9	Q64745	ced-3 (C7020T) gene.	0.00e+00
9	998	99.8	7653	9	Q64742	ced-3 (C6434T) gene.	0.00e+00

10	998	99.8	7653	9	Q64743	ced-3 (C6485T) gene.	0.00e+00
11	998	99.8	7653	9	Q64740	ced-3 (C6322T) gene.	0.00e+00
12	998	99.8	7653	9	Q54666	ced-3 gene.	0.00e+00
13	998	99.8	7653	9	Q64737	ced-3 (G5757A) gene.	0.00e+00
14	298	29.8	2485	14	Q79970	interleukin-1 beta co	2.66e-159
15	127	12.7	7653	9	Q64738	ced-3 (C5940T) gene.	2.76e-56
16	127	12.7	7653	24	T38196	Nematode Ced-3 gene.	2.76e-56
17	127	12.7	7653	9	Q64736	ced-3 (G2487A) gene.	2.76e-56
18	127	12.7	7653	9	Q64745	ced-3 (C7020T) gene.	2.76e-56
19	127	12.7	7653	9	Q64739	ced-3 (G6297A) gene.	2.76e-56
20	127	12.7	7653	9	Q64742	ced-3 (C6434T) gene.	2.76e-56
21	127	12.7	7653	9	Q64743	ced-3 (C6485T) gene.	2.76e-56
22	127	12.7	7653	9	Q64741	ced-3 (G6372A) gene.	2.76e-56
23	127	12.7	7653	9	Q64737	ced-3 (G5757A) gene.	2.76e-56
24	127	12.7	7653	9	Q54666	ced-3 gene.	2.76e-56
25	127	12.7	7653	9	Q64744	ced-3 (G6536A) gene.	2.76e-56
26	127	12.7	7653	9	Q64740	ced-3 (C6322T) gene.	2.76e-56
27	127	12.7	7653	9	Q54401	ced-3 gene.	2.76e-56
28	40	4.0	91	9	Q51746	Oligonucleotide probe	1.44e-07
29	38	3.8	204	1	N81164	Base substituted E.co	1.43e-06
30	37	3.7	91	9	Q51746	Oligonucleotide probe	4.44e-06
31	37	3.7	204	1	N81164	Base substituted E.co	4.44e-06
32	33	3.3	91	46	V44650	Mammalian DNA replica	3.80e-04
33	32	3.2	91	46	V44650	Mammalian DNA replica	1.13e-03
34	30	3.0	114	12	Q70465	Generic DNA sequence	9.59e-03
35	30	3.0	114	12	Q70469	Generic DNA sequence	9.59e-03
36	30	3.0	114	12	Q70468	Generic DNA sequence	9.59e-03
37	30	3.0	501	3	N50025	Sequence encoding new	9.59e-03
38	29	2.9	114	12	Q70467	Generic DNA sequence	2.75e-02
39	28	2.8	498	3	N50034	Sequence encoding new	7.77e-02
40	28	2.8	501	3	N50029	Sequence encoding new	7.77e-02
41	28	2.8	501	3	N50032	Sequence encoding new	7.77e-02
42	28	2.8	501	3	N50027	Sequence encoding new	7.77e-02
43	28	2.8	501	3	N50031	Sequence encoding new	7.77e-02
44	28	2.8	501	3	N50030	Sequence encoding new	7.77e-02
45	28	2.8	501	3	N50024	Sequence encoding new	7.77e-02

ALIGNMENTS

RESULT	1	Q54401 standard; DNA; 7653 BP.
ID	Q54401;	01-JUL-1994 (first entry)
AC	Q54401;	
DT	01-JUL-1994	(first entry)
DE	ced-3 gene.	
KW	C.elegans; cell death; gene; ced-3; inhibition; human; parasite;	
KW	interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;	
KW	inflammatory response; nematode; diagnosis; myocardial infarction;	
KW	stroke; degenerative disease; traumatic brain injury; hypoxia; pest;	
KW	pathogenic infection; hair loss; cancer; autoreactive antibody; ss.	
OS	Caenorhabditis elegans.	
FH	Key	Location/Qualifiers
FT	repeat_unit	1356..1472
FT		/tag= a
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t repeat at 1490-1614"
FT	repeat_unit	1490..1614
FT		/tag= b
FT		/rpt_type= INVERTED
FT		/note= "Inverted w.r.t. repeat at 1356-1472"
FT	exon	2232..2366
FT		/tag= c
FT		/number= Exon_1
FT	mutation	2310
FT		/tag= d
FT		/note= "C>T, from allele n1040"
FT	intron	2367..2429
FT		/tag= e
FT		/number= Intron_1
FT	exon	2430..2575
FT		/tag= f
FT		/number= Exon_2
FT	mutation	2487

/gene="ZK1073.2"
/codon_start=1
/protein_id="CAA92228.1"
/db_xref="PID:e1350957"
/db_xref="PID:g3881512"
/db_xref="GI:3881512"
/db_xref="SPTREMBL:Q23393"
/translation="MPLCFTIQRNSPKCEYTYNCSHTSPHFMTYNRFQGAIIIAV
FHIICSGIFTWLVKCIATERSRIELAEIFLASLCLIAIILLGLKESRLLTG
VLAQTLVGLFLVFLALIMGIKPGILDDPDQNGYNSLEELLIVLEIFISLAVIIE
VMFLTIVLKAYSFTDKYNYTGDRDNV"
complement(26600..31014)
/gene="ZK1073.1"
complement(join(26600..26776,27627..27759,27806..27936,
28287..28364,29476..29562,29764..29999,30879..31014))
/gene="ZK1073.1"
/note="CDNA EST EMBL:D68827 comes from this gene; CDNA EST
EMBL:D74951 comes from this gene; CDNA EST EMBL:D72043
comes from this gene; CDNA EST EMBL:D72143 comes from this
gene; CDNA EST EMBL:C08846 comes from this gene; CDNA EST
yk493b4.3 comes from this gene; CDNA EST yk489g5.3 comes
from this gene; CDNA EST yk489g5.5 comes from this gene;
CDNA EST yk456c2.3 comes from this gene; CDNA EST
yk456c2.5 comes from this gene; CDNA EST yk363b1.3 comes
from this gene; CDNA EST yk363b1.5 comes from this gene;
CDNA EST yk340a9.3 comes from this gene; CDNA EST
yk340a9.5 comes from this gene; CDNA EST yk237g2.5 comes
from this gene; CDNA EST yk330h8.5 comes from this gene;
CDNA EST yk332e12.5 comes from this gene; CDNA EST
yk416g6.5 comes from this gene"

gene
CDS
BASE COUNT 12113 a 6942 c 6586 g 11999 t
ORIGIN
Query Match 8.8%; Score 88; DB 21; Length 37640;
Best Local Similarity 86.6%; Pred. No. 1.75e-35;
Matches 110; Conservative 0; Mismatches 16; Indels 1; Gaps 1;
DB 7713 TAATGGGGTTATTCATTGTGATGACGAAAAATGTAAAAA-GTATTTTATACATTTTCC 7771
|||||
CP 3950 TAATGGGGTTAATCATGTGATGTTGAAAAATACAAAAATGTATTTTAAATACATTTTCCC 3891
DB 7772 CCTATTCATGTGCGAGAAAAATGTTTAAAAAAGCATGATTTTTCGACAT 7831
|||||
CP 3890 CCTATTCATTTGTGCGAGAAAAAGTGAATAAAAAAGCATGATTTTTCGACAT 3831
DB 7832 TTTTCTA 7838
|||||
CP 3830 TTTTCTA 3824

Search completed: Sat Aug 7 18:51:03 1999
Job time : 1848 secs.


```

LRLKYLKLSAFAIEVPPFVLTVRFIATFRASNYENRYKWCGLITGEFHVCLSMFLPSL
HSSGKAGKAIYYCWNVSSTDNQSMRTIPILIVVSOLITIPGLLYLLRKNRERFALSF
OKRSTLSQRYQLSENIQTLTKRFRIVASAVTWFTVYNAATFGYVLFSDAWSYAEERAV
IEIVHCLPLXYLLIAACLIIEDKKHRNNTIKVNNPEQYFNHLQAFDEAFETIVKSR
SRKKIKIPK"
gene
complement(16426. .18477)
/gene="C04F5.6"
complement(join(16426. .16622,17216. .17414,17457. .17564,
17732. .18119,18419. .18477))
/gene="C04F5.6"
/note="contains similarity to C. elegans chemosensory
receptor SRA-10 (GB:Z37092)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69207.1"
/db_xref="PID:g3806134"
/db_xref="GI:3806134"
/translation="MPNETDCQEMAIIVATSSFLRAVLIAVLLCICMIPISYSLMRI
YFSVKLHFNSKIVLTNTFTVLHCLARIVLHGKDLNFPDNWSSGCDILPSRACHV
RFIYRGEYIIIEISPEILLIEREIKADSYENRRKLYGIIFFITHTSTANRLSVN
IPTFIEFTQALIPGLLYLRKNEELRASSLHKHSTLTERYQISENLRTSTMERIMS
VVTWVYAVKAGSYMFIIESVLPDLFAITEVIEHCLPLYIILALATIRVDRKPRS
EFKVNNEYHHFVELHKFEDGAEERKQOVGPMGQKNKIVP"
complement(1816. .21387)
/gene="C04F5.7"
complement(join(1816. .19209,19266. .19591,19641. .19760,
19807. .19954,20000. .20220,21076. .21387))
/gene="C04F5.7"
/note="contains similarity to UDP-glucuronosyl and
UDP-glucosyl transferases (Pfam: UDPGT.hmm, score: 57.39)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69208.1"
/db_xref="PID:g3806135"
/db_xref="GI:3806135"
/translation="MYVSHDFHMRFAFFATFLKHAEARALLTVMDQGRSHATSINT
FMHRLQKDNHTTALFEAFHRODIDGMDERFIDMSGFQNPESPEDKIAFEEDYSFI
HQAIVAGFGSQTNNITLKHRRERFEILYEDMDIYLSLFAICGYMAEISGPKHVM
MHSYDLEAAAGSFKFSRNYATFVPSNLPEFSLNNTVSNYHAWAAYDWEGSFITTA
YVGNFAQKWLRSIIPPIYFSFAEYNNRSSFTETDPSLFPASRTNDFESFGAYCK
ESSKPLDLEKTFIEHPSKSGTILIAFGTIDWRKAPKNYDAFATVNNRLSEYRIIW
SMKGERPGLKKHVTSSWVPQOILHNNKTVLFLSHGLKSTKEVICSATPTIFPM
FGEOTRNAMLKEKGFARIMNFKINDELDTMREVLHPNQQONANKFLTYYMOP
IPTLDEGAFKFNRLVYXGKMPSHFYPKSLDLSYFTVLNVDDLIVPIMMLYLVK"
complement(22210. .24185)
/gene="C04F5.8"
complement(join(22210. .22572,23161. .23248,23629. .23876,
24045. .24185))
/gene="C04F5.8"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69201.1"
/db_xref="PID:g3806128"
/db_xref="GI:3806128"
/translation="MMDRGARKLDRIAAYLRDKAVDVPKAPSKLRWPTEGFSDCTE
LNTCEVDSFLYEDDDVEELVERDELSTRTYCPVCSRDYVPLNFISLSRLQYKFLFD
LMPPTCQHVLDIGSRLGAIYVGASIFTNVSTIGVEMSDYALALRTIRQFGLKN
IEILRDDIRHLAHVFKDNOFIVANNVFSFLSKTEQEECWELHTMRPGTVLAHHP
IEKITEHLDSFEDIEDWLEFVNTTPECLEYGANDSEIYEEMASICKYIVRDSAHGDS
VPE"
complement(25408. .29777)
/gene="C04F5.9"
complement(join(25408. .25796,26398. .26911,28047. .28538,
Note: remainder of annotations omitted.

Query Match 9.48; Score 94; DB 22; Length 39752;
Best Local Similarity 84.28; Pred. No. 2.49e-39;
Matches 128; Conservative 0; Mismatches 22; Indels 2; Gaps 2;

Db 34423 TGCATTTTTCGACATTTTCTACATCGGCCGACATCCGACTTCACATGCTGAAGCGAGA 34482
|||||
OY 3694 TGCATTTTTCGACATTTT-TACATCGAACGACAGCTCAGCTTCACATGCTGAAGCGAGA 3752
|||||

```

```

Db 34483 GACGCGAGAAATGAACCTCCCTCTGCGCTCTCTCGCCTTCAGCATGTGAAGTGAC 34542
|||||
OY 3753 GACGCGAGAAATACACACATCTT-TCGCGCTCTCTCGCTTCAGCATGTGAATGGGA 3811
|||||

Db 34543 TCTCGCTCGATGTGAATAATGTGCAAAATG 34574
|||||
OY 3812 TCTCGCTCGATGTGAATAATGTGCAATAATG 3843
|||||

RESULT 15
LOCUS CEZK1073 37640 bp DNA INV 23-NOV-1998
DEFINITION Caenorhabditis elegans cosmid ZK1073, complete sequence.
ACCESSION Z68135
NID 91753051
VERSION Z68135.1 GI:1753051
KEYWORDS HTG; Transposon.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
McMurray, A.
REFERENCE 1 (bases 1 to 37640)
AUTHORS Direct Submission
TITLE Submitted (30-NOV-1995) Louis, MO 63110, USA. E-mail:
jess@sanger.ac.uk or tw@nematode.wustl.edu
2 (bases 1 to 37640)
REFERENCE 2
AUTHORS Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, M., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,
Saunders, D., Showkhen, R., Smaildon, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Spratt, J., and Woldman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
Nature 368 (6466), 32-38 (1994)
94150718
JOURNAL On Dec 25, 1996 this sequence version replaced gi:1100866.
MEDLINE Coding sequences below are predicted from computer analysis, using
COMMENT predictions from GeneFinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:-
http://webace.sanger.ac.uk/cgi-
bin/display?db=wormace&class=Sequence &object=ZK1073
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
This sequence is the entire insert of clone ZK1073. The true left
end of clone F09C8 is at 30045 in this sequence. The start of this
sequence (1..104) overlaps with the end of sequence Z70681.
The end of this sequence (37537..37640) overlaps with the start of
sequence Z68132.

FEATURES
location/Qualifiers
source 1..37640
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="X"
/clone="ZK1073"
Z3268. .25951
/gene="ZK1073.2"
join(23268. .23585,25170. .25272,25803. .25951)

```


3775 GATGTGTGG-TATTTCTCCGCTCTCTGCTTCAGCATGTGAAGTGTGCTTGA 3717
34553 TGTGAAAAATGTCGAAAAATGCA 34576
3716 TGTGAAAAATGTCGAAAAATGCA 3694

RESULT 14
LOCUS CELC04F5 39752 bp DNA INV 29-OCT-1998
DEFINITION Caenorhabditis elegans cosmid C04F5.
ACCESSION AF101305
NID 93806127
VERSION AF101305.1 GI:3806127
KEYWORDS
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
REFERENCE 1 (bases 1 to 39752)
AUTHORS Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A., Saunders, D., Showkneen, R., Smaildon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J. and Wohlman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans

JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 39752)
AUTHORS Latreille, P., Bradshaw, H. and Elliott, G.
TITLE The sequence of C. elegans cosmid C04F5
JOURNAL Unpublished (1998)
REFERENCE 3 (bases 1 to 39752)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: rw@nematode.wustl.edu and jesus@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' clone is Y58A7A, 200 bp overlap; 3' clone is F14F9, 2900 bp overlap. Actual start of this clone is at base position 1 of CELC04F5; actual end is at 38256 of CELC04F5

NOTES:

FEATURES

source

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).

Location/Qualifiers

1. 39752

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/clone="C04F5"

/chromosome="V"

/complement(4336..6898)

/gene="C04F5.2"

complement(join(4336..4484,5076..5296,5357..5472,5544..5707,6181..6273,6391..6560,6612..6748,6815..6898))

/gene="C04F5.2"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC69202.1"

/db_xref="pid:93806129"

/db_xref="GI:3806129"

/translation="MYSVHATSVMASSSINCTSVMLRNHFIVAVVDKIAQISGLY
ALLIVESSLFFLLTVHFKIRGSHPEFGLLFGVLLCYICGVAVCOFLFVG
SSDLSVYLYIDLAYAATTAIVFCCLDLRNCRKIRDIVFCKLYCRKERSQVFLIV
YKRTIATLMAVSIEYVTRTWLFLISVCLSVGYIYFASFRCRCIYSAVVTIIIGGC
IALYFLNRKWTILYOKDISFMSRYSRYOIEEMISVRKALLPYIILDSLVTVDA
LSSKVTTKIEEQSNAMCYSMNTYIPYIILNLMAVEYSIPLYMFRKAYSKKLPNVC
HRIPIARIQOSRESMIONVEGKELNVVPTQAEYFAKLQNEWK"

complement(7985..9484)

/gene="C04F5.3"

complement(join(7985..8175,8663..8727,8773..8887,8946..9021,9072..9197,9267..9341,9449..9484))

/gene="C04F5.3"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC69204.1"

/db_xref="pid:93806131"

/db_xref="GI:3806131"

/translation="MLSQLPLQLQAPGDTCSIVTFQVKILNQNNNSVPIDFHLPSGL
IKGNCVTRKSKAVISSTIEETDGRMTLKFFRTEEMRAKRLDELRLWLNKVIYEK
FAGSAVFEESDSSVIFSAPLQKYCEDRINVTLSVNFDPILVMFSPEIDVQPYG
PKSNFYICERSRKRTLSESLQHRSTVFCGVVALLSLAHIVGMLRRHFMHARKELYE
NLDKLPAMN"

complement(10719..12738)

/gene="C04F5.4"

complement(join(10719..10912,10970..11171,11228..11335,11385..11459,11875..12262,12680..12738))

/gene="C04F5.4"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC69205.1"

/db_xref="pid:93806132"

/db_xref="GI:3806132"

/translation="MPNETDCQVMAEVAATSSFLRTSLGNTLLCICLVCPTIYSFLAI
OKATKLFHNSKCIFFHTLTFAFIMIVRIILHGKDLNYPGPMWNGCEIFPSRSRCEL
RLYKISAFVEVEVTEPVLTAERFVATFERARHYENRYKWCGLNIFHISLALFLFTIQ
SSEKVGDIYYCWMSSSTGNRYMLNLPFVIVFSQLTIPALLYLLRKNEKFRASIQ
KRSTLSQRYQLSQNLQTLTKFRIVSTVTWYVYVYNAATFGVHFFLKSSMSAGQFAII
EIVHCLPLYLLIFLWVKEDKKPHROFSIKVDHYEPQYFNDLQKFFDQSFDKVKKTK
SVTFVVS"

complement(13663..15655)

/gene="C04F5.5"

complement(join(13663..13856,14037..14238,14292..14399,14454..14528,14601..14988,15594..15655))

/gene="C04F5.5"

/note="contains similarity to C. elegans chemosensory receptor SRA-10 (GB:Z37092)"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC69206.1"

/db_xref="pid:93806133"

/db_xref="GI:3806133"

/translation="MSINETDCQIMAEIATSSFLRTVLSTTFLCICICPCLNFAWL
IHETRLHNSKCIFFVTFNISVLYVHVSRIILHGKDFLNYVGPWNGCELLFPDRSRCD

gene
complement(4336. .6898)
/gene="C04F5.2"
complement(join(4336. .4484,5076. .5296,5357. .5472,
5544. .5707,6181. .6273,6391. .6560,6612. .6748,6815. .6898))
/gene="C04F5.2"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69202.1"
/db_xref="PID:g3806129"
/db_xref="GI:3806129"
/translation="MYSVHATSVMASSSLNCTSVMLRNHFIYAVNVDKIAISIGLY
ALLLVESLFFLLTVHFKIRGSHPEFGLGCVLLCYIICGVAEAVCOFLFEG
SSDLSVYLLIDLAYAATTAIVFCCLIDILNCRKIRDIIFCKLYCRKERSOVFLIY
YKRTIATLMAVSLEYRTWLEFLISVCLSVGYIYFASFRCRCIYSAVVTIIIGGC
IALYFLNRKWTILYQKDISFMGKVSLSMRQIIEIMISVRKALLPVIIDSLVTLVDA
LSSKVTTKIEEOSNACYSMTNIPYILNHLAMVEYSIPLYMFRKAYSKKLPNV
HRIPIARIQOSRESMIQNVGKELNVPTQAEYFAKLQNEWK"
complement(7985. .9484)
/gene="C04F5.3"
complement(join(7985. .8175,8663. .8727,8773. .8887,
8946. .9021,9072. .9197,9267. .9341,9449. .9484))
/gene="C04F5.3"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69204.1"
/db_xref="PID:g3806131"
/db_xref="GI:3806131"
/translation="MLSQPLQLQAPGDTICISVTFQVKILNOONNSVPIDFHLPSGL
IKGNCAVTRKSKAVISSTIEETDGRMKTLEKVERTEEMRAKRLDELRLQNLKVIYEK
FAGSSAVFESDSSSVIFSAPLQKYCEDRINVTLSVNFDPILVMFSPEDVOPYG
PKSNFYICERSRKRTLSLSQHRSTVFCGVVLLASSIAHYVGMRLRRHFMHRKELYE
NLDKLPAMN"
complement(10719. .12738)
/gene="C04F5.4"
complement(join(10719. .10912,10970. .11171,11228. .11335,
11385. .11459,11875. .12262,12680. .12738))
/gene="C04F5.4"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69205.1"
/db_xref="PID:g3806132"
/db_xref="GI:3806132"
/translation="MPNETDCOVMAEVATSSFLRTSLGNTLLCICVPCITYSFLAI
OKARLHNSKCIHHTLTFATFIMIVRIILHGKDLNVGPMWGCETIPSPSRCEL
RKLYKISAFVEVTPFVLTAERFAVTERARHYENRYKWCGLNIFHISLAFLEFTIO
SSEKYGDDIYYCWMSSGTGRMLNLPFIVIVFSQITIPALYLLRKNEKPREASLO
KRSTLSQRYQLSOLQTLTKFRIVSTVTWIVTYNAATFEGVHFLKSMSSAQFOAI
EIVHCLPLYLLILFLMWKEDKKPHROFSIKVDHYEPQYFNDLQKFQDSFQKVKKT
SVTFVVS"
complement(13663. .15655)
/gene="C04F5.5"
complement(join(13663. .13856,14037. .14238,14292. .14399,
14454. .14528,14601. .14988,15594. .15655))
/gene="C04F5.5"
/note="contains similarity to C. elegans chemosensory
receptor SRA-10 (GB:Z37092)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69206.1"
/db_xref="PID:g3806133"
/db_xref="GI:3806133"
/translation="MSINETDCOIMAEIATSSFLRTVLTSTTFCLICICPFCNLFAWL
IIHETRLHNSKCIYFTNISVLVHVSRIILHGKDLNVGPMWGCETIPDRSRCD
LRKLYKISAFIEVTPFVLTAERFAVTERARHYENRYKWCGLLTHGVCLSMFLFSL
HSSGKAGKAIITYCWVSTNDQSMRTIPILIVVSQITIPGLVLLRKNERFALSF
OKRSTLSQRYQLSOLQTLTKFRIVSAVTVTYNAATFEGVHFLSDAWSYAEFAV
EIVHCLPLYLLILAACLIEDKKHRNQNIIKVNNEYEPQYFNHLQAFDEAFETVKS
SRKVKIPK"
complement(16426. .18477)
/gene="C04F5.6"
complement(join(16426. .16622,17216. .17414,17457. .17564,
17732. .18119,18419. .18477))
/gene="C04F5.6"

gene
/note="contains similarity to C. elegans chemosensory
receptor SRA-10 (GB:Z37092)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69207.1"
/db_xref="PID:g3806134"
/db_xref="GI:3806134"
/translation="MPNETDCOEMAIVATSSFLRAVLLIALLCIMCIPISISLWRI
YFSVKLHNSKIVLFTNTEFVLHICLARIVLHGKDLNVFDMWSSGCDLPSRACHV
RFIYRGEYIIIEISPFILIERFIATFKADSYENRKLGIIFITHTSISANRLSVN
IPTEFIFPTQATIPGLLYLRKNEELRASSLHKSTLTERYQISENLTSTMRMS
VTVWVYVAYKAIGSYMXYFIESVLLPDLFAITEYIHCLPLYIILALITRVDKPKRS
EFKVNNEYTHFVELHKEFDGAFERKQQVPMQKNKIVP"
complement(18816. .21387)
/gene="C04F5.7"
complement(join(18816. .19209,19266. .19591,19641. .19760,
19807. .19954,20000. .20220,21076. .21387))
/gene="C04F5.7"
/note="contains similarity to UDP-glucosyl and
UDP-glucosyl transferases (Pfam: UDPGT.hmm, score: 57.39)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69208.1"
/db_xref="PID:g3806135"
/db_xref="GI:3806135"
/translation="MYVSHDFHMRIFAEATFLKHAEPARALLTVMDGRSHATSINT
FMHRLQKDNHTALEFAHAFRODIDFGMDERFIDMSGFQNPESDFDKIAFEEDYSFI
HOAVAYFGSGQTCNNILKRRERFEILYEDWDIYLSDSLFAICGYMAEISGPHVM
MHSTDLEAAQSGSEKFSRNRYATFVPSNLPFSMLNTVSNYIHRWAAIYDWFSGFI
YVGNFAQKWLARSIIPEFYFSAEYVNRSSFTFTDMPDSLFPASRTNDFSGAYCK
ESSKPLDLEKTFEIEHPSKGTILAFGTIFIDWRKAPKNYDAFATVNRLESEYRIIW
SMKGERPGLKHKVKTSSWVPQNLHNNKTVLFLSHGSKREVICSATPTIFVPM
FGEOTRNAMLKEKGFARIMNKEKINDELDTMREVLHPNQONANKFLTYYMDQ
IPTLDEGAFENRLVYGGKMPSHFYPKSLDSLSTFTLVNVDLIVPIMMLYLVK"
complement(22210. .24185)
/gene="C04F5.8"
complement(join(22210. .22572,23161. .23248,23629. .23876,
24045. .24185))
/gene="C04F5.8"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69201.1"
/db_xref="PID:g3806128"
/db_xref="GI:3806128"
/translation="MMDRGARKLDRIAAYLRDKAVDVPSKASKLWPTTEGDSCTE
LNTCEVDSFLYEDDDVEELVERDELSTRTYCPVCSRRPTVPLNFIHSLSLRQVFLFD
LMPPTCOHVLIDIGSRIGAIYVYASIFITNGNVSTIGVPMDSYALALRTIRQFLKN
IEILRDDRHLAHVFKDNQFIVMNVVFSFLSKTEQECWEFLHTHMRPGTVLAHNP
IEKITEHLDSLFDIEDWLEFVNTTECLEYGANDSEIYEEMASICKYIVRDSRAHGDS
VPE"
complement(25408. .29777)
/gene="C04F5.9"
complement(join(25408. .25796,26398. .26911,28047. .28538,
29777))

Note: remainder of annotations omitted.

Query Match	10.5%	Score 105;	DB 22;	length 39752;
Best Local Similarity	78.4%	Pred. No. 1.84e-46;		
Matches	207;	Conservative 0;	Mismatches 48;	Indels 9; Gaps 6;
Db 34313	ATGAGGTTATTCATGGCTGACGTAATAATGCAAAAATTCATTTTATACAAATGCT	34372		
Cp 3948	ATGGGGTTAATCATGTGATGTTGAATAATACAAAAATG- TATTTAATACATTT-TC	3892		
Db 34373	CCCTATTACGTAATGCAAAAAAATGTTTGAATACATGCAATTTTTCATTTTC	34432		
Cp 3891	CCCTATTCAATTGTCAGAAAAG--TGTAAAAA-ACGCATGCAATTTTTCATTTATTC	3836		
Db 34433	GACATTTTCTACATCGCCGACATCCGACTTCACATGCTGAAGCGAGAGCCGAGAGA	34492		
Cp 3835	GACATTTTCTACATCGACCGAGATCCCATTTTCACATGCTGAAGAGAGAGCCGAGAAA	3776		
Db 34493	AATAAAAACTCCCTCTCTGCGTCTCGCCTTCACGATGTGAAGTGAGCTGTGCTCGA	34552		


```

/ gene="ced-3"
/ number=2
exon      2854. .3107
/ gene="ced-3"
/ number=3
intron    3108. .4302
/ gene="ced-3"
/ number=3
exon      4303. .4634
/ gene="ced-3"
/ number=4
intron    4635. .5546
/ gene="ced-3"
/ number=4
exon      5547. .5760
/ gene="ced-3"
/ number=5
intron    5761. .5814
/ gene="ced-3"
/ number=5
exon      5815. .5942
/ gene="ced-3"
/ number=6
intron    5943. .6297
/ gene="ced-3"
/ number=6
exon      6298. .6537
/ gene="ced-3"
/ number=7
intron    6538. .7012
/ gene="ced-3"
/ number=7
exon      7013. .7652
/ gene="ced-3"
/ number=8
BASE COUNT 2429 a 1455 c 1271 g 2498 t
ORIGIN

```

Query Match 12.7%; Score 127; DB 21; Length 7653;
 Best Local Similarity 78.0%; Pred. No. 5.87e-61;
 Matches 309; Conservative 0; Mismatches 68; Indels 19; Gaps 13;

```

Db 3463 GTTCGCAGAT-TTTTGATTTTTCATTCAGATATGCTTATTAACATATATTAT 3521
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cp 4087 GTTCGCACATATTTTCTGCTGACTTTCATTCAGATATGCTTATTAACATATATTAT 4028
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3522 CATTAATGTAATTTCTGTAGAAATTTGGCTTTTCTAGTACTTCTACTTTTG 3581
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cp 4027 CAATAAAGTAATATCTTGAAGAAATTTTG--TTTCGCTTAATATTTCTACTTTT 3970
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3582 AAATGCTCAACGAAATAATCATGTGTTGTTCA-TATGAATGACGAAATAAGCAAT- 3639
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cp 3969 AAGTTGCTCAACGAAATAATGCGGTTAATCATGTGATGTTGAAATAATACAAAAATG 3910
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3640 T-TTTTA-TATATTTTCCC-TATTCATGTTGTCAGAAAAATAGTAAAAA-GC--GCA 3693
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cp 3909 TATTTTATACATTTTCCCTATTCAT-TTGTGAGAAAAAGTAAAAAAACGCATGCA 3851
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3694 TG-----CATTTTCGACATTTT-TACATCGAAGACAGACTTCACTTCACATGCTGAAGA 3747
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cp 3850 TTTTTCACATTAATTCGACATTTTTCATCGACGAGATCCCATTTTCACATGCTGAAGA 3791
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3748 CGAGAGACGCGGAGAAATACACACATCTTCTGCGTCTCTCTTCAGCATGTGAAT 3807
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cp 3790 CGAGAGACGCGGAGAAAGATGTGTATTTCTCCGCTCTCTCTTCAGCATGTGAAGT 3731
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3808 GGGATCTCGGTCGATGTAAATAATGTCGAATAATG 3843
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cp 3730 GAGCTGCTGTCGATGTAAAAA-TGTCGAAAAATG 3696
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 13 CELC04F5 39752 bp DNA INV 29-OCT-1998
 LOCUS

DEFINITION Caenorhabditis elegans cosmid C04F5.
 ACCESSION AF101305
 NID 93806127
 VERSION AF101305.1 GI:3806127
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Caenorhabditis elegans cosmid C04F5.
 AF101305
 93806127
 AF101305.1 GI:3806127
 Caenorhabditis elegans.
 Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 39752)
 Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
 Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
 Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
 Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
 Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
 Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
 O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A.,
 Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E.,
 Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
 Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
 Wilkinson-Sproat, J. and Woldman, P.
 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans
 Nature 368 (6466), 32-38 (1994)
 94150718
 2 (bases 1 to 39752)
 Latreille, P., Bradshaw, H. and Elliott, G.
 The sequence of C. elegans cosmid C04F5
 Unpublished (1998)
 3 (bases 1 to 39752)
 Waterston, R.
 Direct Submission
 Submitted (26-OCT-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 e-mail: rw@nematoe.wustl.edu and jesus@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate
 chemistry; an attempt was made to resolve all sequencing problems,
 such as compressions and repeats; all regions were covered by
 sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' clone is Y58A7A, 200 bp overlap; 3' clone is F14F9, 2900 bp
 overlap. Actual start of this clone is at base position 1 of
 CELC04F5; actual end is at 38256 of CELC04F5

NOTES:

Coding sequences below are predicted from computer analysis, using
 the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES
 source
 Location/Qualifiers
 1..39752
 /organism="Caenorhabditis elegans"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /clone="C04F5"
 /chromosome="V"

Query Match	14.18;	Score 141;	DB 19;	Length 110000;
Best Local Similarity	82.0%;	Pred. No. 2.62e-70;		
Matches 219;	Conservative	0;	Mismatches 42;	Indels 6;
				Gaps 5;
Db 44711	GAAGCATATATGGGGTTAATCATGTGATGACGAAAAATGCAAAA-TGTATTTTCATACA	44769		
Cp 3957	GAATAAATAATATGGGGTTAATCATGTGATGTTGAAAAATACAAAAAATGTATTTAATACA	3898		
Db 44770	TTTTCCCCCTATTCATGTGTAGAAAAATGTTTTTAAAAAAACGCATGCATTTTGACA	44829		
Cp 3897	TTTT-CCCCCTATTCATTTGTGTGAGAAAAAGTGA-AAAAAACGCATGCATTTTTT-ACA	3842		
Db 44830	TTTTTCTACATTTTTTCTACATCGACCGACAGTCCACTTCTCATGCTGAAGACAGAGACG	44889		
Cp 3841	TTATTCGACATTTTTTTTACATCGACCGACAGATCCCATTTTCACATGCTGAAGACAGAGACG	3782		
Db 44890	CAGAGAAATACACACACATCTCTGCGTCTCTCGCCTTCAGCATTTAAAGTGAGCTGTG	44949		
Cp 3781	CAGAAAGATGTGTGTATTTCTCCGCGTCTCTGCTTTCAGCATGTGAAGTGAGCTGTG	3722		
Db 44950	GCCGATGTAGAAAAATTTCGAAAAATG 44976			
Cp 3721	TTTCGATGTA-AAAAATGTCGAAAAATG 3696			

RESULT	11			
LOCUS				
DEFINITION	CEY62E10	256941 bp	DNA	HTG 12-MAR-1999
	Caenorhabditis elegans	DNA *** SEQUENCING IN PROGRESS ***		from
	clone Y62E10,	WORKING DRAFT SEQUENCE.		
ACCESSION	AL031580			
NID	94469033			
VERSION	AL031580.4	GI:4469033		
KEYWORDS	HTG; HTGS_PHASE1.			
SOURCE	Caenorhabditis elegans.			
ORGANISM	Caenorhabditis elegans.			
	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;			
	Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.			
REFERENCE	1 (bases 1 to 256941)			
AUTHORS	Lloyd, C.			
TITLE	Direct Submission			
JOURNAL	Submitted (12-MAR-1999)	Nematode Sequencing Project, Sanger Centre.		

COMMENT

On Mar 22, 1999 this sequence version replaced gi:4455386.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.

- * NOTE: This is a 'working draft' sequence.
- * This record will be updated with the finished sequence
- * as soon as it is available and the accession number will
- * be preserved.

FEATURES	Location/Qualifiers
source	1. .256941
	/organism="Caenorhabditis elegans"
	/db_xref="taxon:6239"
	/chromosome="IV"
	/clone="Y62E10"
BASE COUNT	77397 a 44556 c 43039 g 79104 t 12845 others
ORIGIN	

Query Match	14.18;	Score 141;	DB 19;	Length 256941;
Best Local Similarity	82.08;	Pred. No. 2.62e-70;		
Matches	219;	Mismatches 42;	Indels 6;	Gaps 5;

Dd 112946 GAAGATATATGGGTTAATCATGTGATGACGAATAATGCAAAA-TGTATTTCATACA 113004
||||| - ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cp 3957 GAAAAAATAATGGGTTAATCATGTGATGTTGAAAAATACAAAAATGTATTTTAATACA 3898

Db 113005 TTTTCCCCCATTCAATGTCGTAGAAAAAATGTTTTTTTAAAAAAACCGCATGCACCTTTTCGACA 113064
 |||| ||||| ||||| ||| | ||| || | ||||| ||||| ||||| ||
 Cp 3897 TTTT-CCCCCATTCAATTTGTGCGAAAAAGTGTA- -AAAAAACCGCATGCAATTTTTT-ACA 3842
 Db 113065 TTTTCTACATTTTCTACATCGACCGACAGTCCACTTCTCATGCTGGAAGACGAGAGACG 113124
 || ||| ||||| ||||| ||||| || ||| ||| ||||| ||||| ||||| ||
 Cp 3841 TTATTCGACATTTTCTTACATCGACCGAGATCCCATTTTCACATGCTGGAAGACGAGAGACG 3782
 Db 113125 CAGAGAAATACCACACACTCTCTGCGTCTCTCGCCTTCAGCATTTAAAGTGAGCTGCG 113184
 |||| | || || ||||| ||||| ||||| ||||| || ||||| ||||| ||
 Cp 3781 CAGAAAGATGTGTGTATTTCTCCGCGCTCTCTCGTCTTCAGCATGTGAAGTGAGCTGCG 3722
 Db 113185 GCCGATGTAGAAAAAATTTCGAAAAATG 113211
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
 Cp 3721 TTTCGATGTA-AAAAATGTCGAAAAAATG 3696

RESULT	12
LOCUS	CeLCED3A 7653 bp DNA INV 23-FEB-1994
DEFINITION	Caenorhabditis elegans cell death protein (ced-3) gene, complete cds.
ACCESSION	L29052
NID	g456416
VERSION	L29052.1 GI:456416
KEYWORDS	cell death protein; interleukin-1 beta converting enzyme.
SOURCE	Caenorhabditis elegans (strain N2) DNA.
ORGANISM	Caenorhabditis elegans Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda; Secernentea; Rhabdilitia; Rhabditida; Rhabditoidea;
REFERENCE	Rhabditidae; Caenorhabditis. 1 (bases 1 to 7653)
AUTHORS	Yuan,J., Shaham,S., Ledoux,S., Ellis,H.M. and Horvitz,H.
TITLE	The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1b-converting enzyme
JOURNAL	Cell 75, 641-652 (1993)
MEDLINE	94061982

FEATURES	Location/Qualifiers
source	1..7653 /organism="Caenorhabditis elegans" /strain="N2" /db_xref="taxon:6239" 2167..2366 /gene="ced-3" /number=1
gene	join(2232..2366,2430..2575,2854..3107,4303..4634,5547..5760,5815..5942,6298..6537,7013..7075) /gene="ced-3"
CDS	join(2232..2366,2430..2575,2854..3107,4303..4634,5547..5760,5815..5942,6298..6537,7013..7075) /gene="ced-3" /note="codes for a protein similar to mammalian interleukin-1b-converting enzyme" /codon_start=1 /product="cell death protein" /protein_id="AA027982.1" /db_xref="PID:q456417" /db_xref="GI:456417"
intron	/translation="MMRQDRRLLEERNIMFSSHLKVDLEILEVLAQVLSNDGMDMI NSGTVREKREIVKAVQRRGDVAFDAFYDALRSTGHEGLAEVLEPLASVDSNAVEF ECPMPASHRRSRALSPAGYTSPTRVHRDSVSSVSTSYQDIYSRAERSRSLAHS SDRHYSSPPVNAFPSPQSSANSSTFGCSLGYSSSRNRSFSKASGPTQYIFHEEDMNN FVDAPTISRVEDEKTMTRNPSPRGMCLINNEHEFQMPTRNGTKADKNLTNLFRCM GYTVICKDNLTRGMLTTRIDFAKHESHGSAIILVLSHGEENVILGYDDIPISTHEI YDLNANAPRLANKPKIYFVQACRGERDNGFPLVDSVDGVPAFLRGWDRDGPFLF NFLGCVRPOVOQVWRKKPSQADILLIRYATTAQYVSWRNSARGSWFIQAVCEVFSTHAK MDVVELLTEVNNKKVACGFQTSQGSNILKQMPENTSRLKKFYFWEARNSAV"
exon	2367..2429 /gene="ced-3" /number=1 2430..2575 /gene="ced-3" /number=2
intron	2576..2853


```

YK210a2.3 comes from this gene; cDNA EST yk216a4.3 comes
from this gene; cDNA EST yk216a4.5 comes from this gene;
cDNA EST yk226f6.5 comes from this gene; cDNA EST
yk338a3.3 comes from this gene; cDNA EST yk338a3.5 comes
from this gene; cDNA EST yk351f3.5 comes from this gene;
cDNA EST yk406e1.5 comes from this gene; cDNA EST
yk408e8.3 comes from this gene; cDNA EST yk408e8.5 comes
from this gene; cDNA EST yk427f8.5 comes from this gene;
cDNA EST yk480b8.3 comes from this gene; cDNA EST
EMBL:T00333 comes from this gene"
/codon_start=1
/protein_id="CAB05207.1"
/db_xref="PID:e1354464"
/db_xref="GI:3924776"
/db_xref="GI:3924776"
/translation="MTEDGVKMDASTSTIKEPEPADCVILPTHTNEOVVYKSWIAF
GDNPNLYNRDLGADWGYIVDCSDSPPTGPGVYKAFIYNDLNYQESGNPLFKYTG
RLKFLANKDTSMTMKLEIFEONRLELRSPKMSDSOYLAILEDNETNSAICYS
EQLSSIKYHLPGVSPAKAGQILFTEYNYENDYFPDPRDAIEYRECEMTRIENGE
LLISAKACSSIEITHITPQSAFAEDTGVNVAEEMERGYYRVELAINIIDSEAN
NVPLFKISDIEGKLTGEENAEFLRKNDEELQGTWEDWTEPEIAVELEPVAABEPA
DPEVEAKIEENGYNDESLTQSESDSVFLPQIESASSEEMSENSAENSEKPKTF
GRMPSLSESSSSSTISSKVGDSVIVLVNDNDECYGVETQDKMTICTAGVTVPL
RPGYIFQFIACNNYKTSIQTOSARVSECEGAVFSSDDVPKISANVIFSSNPKHXS
NKEGVSDRYGFEVIEPRITGDTGTVYSTQITMAVQYGDPLACFKAVALPKRQFID
LDHYLOVMKYEDNITQLSRKNGLGAKSIRLPNEPRNGSKHSTKRTTTPHTGLVY
DMEKRGNYFCGWTCFQDVFCEPSONGYPAGVDLVNIEYELISPTKDKDNGI
ASMTKEPASTTIVWRNNGEIVKATILMKPEKLIYSEDNREYEHQOLGIVADDSI
YGSNDEQSKVEFERLRNPITIGFTNLMNKCWYARLPMIHREPSPEPINKYNDEN
VYDDFOKGGPQSVNSYRSRDROYQEKPMQYSHDPAOTISHSQDNHRPMETRSNAG
TSVSSVDREKGYVMKKVDKNTQIVLYSGQAAAIHWNTHSQSTIVGNSYIFMVRPI
EVHKAHYAPEITQVEILNTEHNGVDRSEFSLFTDLCNDQRYVERLAFGEAL
VLRNDVGVVMAASNVKTPPLNPSMAVEYCOQYKGRSESNVGYCRLVYGVKPREIY
DGHFSDQFVSAYYFELHEILATSNAYENMLDRNLRGNTIEVRDIDISQKRIAH
MNFQNEQOQOQSSLSRGPTRPYSPVGSYQSSHQNFDRSAHNGDGRSTQSHHSA
GGSNMHONFOPPYQSSSSQOPRYNGAPSGPLPPSSQSSFNAPAPPHYSEMS
NDYNRMVRKLAEITDTAAQMIINNDIRQIKNISPDLLDLERNIDACQREKRHYEK
YGN"
complement(27575..32349)
/gene="F52G2.3"
complement(join(27575..27979,29638..29814,29857..29959,
30010..30132,30177..30380,30607..30741,31300..31446,
31498..31803,32240..32349))
/gene="F52G2.3"
/codon_start=1
/protein_id="CAB05205.1"
/db_xref="PID:e1354462"
/db_xref="GI:3924775"
/db_xref="GI:3924775"
/db_xref="SPIREMBL:O62256"
/translation="METRLVLKYDILNREALICLNRETHELHANLVESVQOIGOLF
AVENLKIIVQIDGVCFTSPEGCTIVIEPPAPEDCVILPEKTDVVMKSWIGSSDSPH
LTYDRNIGADWGPVNCOSQKAKNTKVAFFKYLIAVNDLDYKNTGCLMKAVOLGVY
IDNRDAEVRLLKMEFERKRIADIQTQRTQRLIETVAYPIPLSAHQSSNHSYTGRS
SCDSLRPQVPARKLSQOPQVONQAPYREQINOQRSVTNSIYGEFSGVGRDQPLE
TANSRGSTEYSGRSSSSDOVIRNEPQLPNPAPILPESMGVTEYENGIMQIVLYRSV
QKAIHYWSKHQITTSLSGVNCSFKPIPVQATTHWAPFEITAINNVNNEKNEVKTG
KECFALFKVDLSRPEGKFTWTIGNGPIFSCETIVGYMLAHSIGDADNVKRHNDI
EMQKKRAQOEHAQOQRPAGSSNSYTENAQOHSMSNMQYHSQAPVSSRSSISLG
ETAEFKNKLERTNDMWSLCAIKKFLDNRLQKETLRYGNPRLQALEKLEKCESEHL
ENN"
complement(33889..34904)
/gene="F52G2.4"
complement(join(33889..33909,33958..34013,34063..34141,
34714..34802,34853..34904))
/gene="F52G2.4"
/notes="cDNA EST yk356f5.5 comes from this gene; cDNA EST
EMBL:D64203 comes from this gene; cDNA EST EMBL:D66267
comes from this gene; cDNA EST yk298f3.3 comes from this
gene; cDNA EST yk356f5.3 comes from this gene; cDNA EST
yk383e1.3 comes from this gene"
/codon_start=1
/protein_id="CAB05208.1"

```

```

/db_xref="PID:e1354465"
/db_xref="PID:g3924777"
/db_xref="GI:3924777"
/translation="MRGVEENDGLVMTMCRVVEKKKDNMVKRDESGRLDEKHN
TWKGLDPLLRDGPNEGHRWSRTVEKMTGDTRFQDVNRQYNKEYVESIVRNY"
BASE COUNT 12956 a 6666 c 6504 g 13145 t
ORIGIN

```

```

Query Match 14.1%; Score 141; DB 21; Length 39271;
Best Local Similarity 82.0%; Pred. No. 2.62e-70;
Matches 219; Conservative 0; Mismatches 42; Indels 6; Gaps 5;

```

```

Db 3110 GAAGATATATGGGTTAAATCATGTGATGACGAAATAATGCAAAA-TGTAATTCATACA 3168
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 3957 GAAAAATAATGGGTTAAATCATGTGATGAAATAATACAAAAATGTAATTAATACA 3898
Db 3169 TTTTCCCCCTATCATGTGTAGAAAATGTTTAAAAAACGATGCACATTTCGACA 3228
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 3897 TTTT-CCCCCTATCATTTGTGACAGAAAAGTGA--AAAAACGATGCATTTT-ACA 3842
Db 3229 TTTTCTACATTTTCTACATGACCGACAGTCCACTTCTCATGCTGAAGACGAGAGACG 3288
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 3841 TTATTCGACATTTTCTACATGACCGAGATCCCATTTTCACATGCTGAAGACGAGAGACG 3782
Db 3289 CAGAGAAATACACACATCTCTGCGTCTCGCCTTCACGATTTAAAGTGAGCTGTCG 3348
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 3781 CAGAAAGATGTGTGTATTCTCCGCGTCTCTCGTTCACGATGTGAAGTGAGCTGTCG 3722
Db 3349 GCCGATGTAGAAAATTTGCAAAAATG 3375
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 3721 TTGCATGTA-AAAAATGTGCAAAAATG 3696

```

```

RESULT 10 CEY45F10 434238 bp DNA HTG 03-DEC-1998
LOCUS Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
DEFINITION clone Y45F10, WORKING DRAFT SEQUENCE.
ACCESSION Z93245
NID 93378089
VERSION Z93245.1 GI:3378089
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 434238)
AUTHORS McMurray,A.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwenematode.wustl.edu

```

```

COMMENT
Order of segments is not known; 800 n's separate segments.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.

```

```

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

FEATURES
source Location/Qualifiers
1..434238

```

```

/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="Y45F10"
BASE COUNT 141590 a 73465 c 71957 g 143994 t 3232 others
ORIGIN

```


Matches	320;	Conservative	0;	Mismatches	52;	Indels	21;	Gaps	9;
Db 239176	CATTATTCGACATTTT	TACATCGACCGACATCCCATTTCA	CATGCTGAAGACGAGAGA	239235					
QY 3696	CATTTTTCGACATTTT	T-ACATCGAACGACAGCTCACCTTCACATGCTGAAGACGAGAGA	3754						
Db 239236	CGCAGAAAGATGTGTG	TATTTCTCCGCTCTCTGCTTCACGATGTGAAGTGAGCTGT	239295						
QY 3755	CGCGAGAAATATACCACACACATCTTCTGCGTCTCTCGTTCAGCATGTGAATGGGATCT	3814							
Db 239296	CGTTCGATGTAAAAAA	-TGTCCGAA-----AAATGCATGCGCTTTT	TACATTTTC	239345					
QY 3815	CGGTGCGATGTAAAAAAATGTCCGAAATATGTAAAAAAATGCATGCGTTT	TTTTCACATTTTC	3874						
Db 239346	TGCACACATGAATAGGGG	-AAAAATATAT-AAAA-A-AT-----GCATTTTTCGTATCAT	239396						
QY 3875	TGCACAAATGAATAGGGGGG	AAAAATGTATTAATAATACATTTT	TGTATTTTCAACATCAC	3934					
Db 239397	ATGAACAACCCACATGATTTT	TTCGTTGAGCAATTTCAAAAGTAGAGCATACTAGAACGA	239456						
QY 3935	ATGATTAACCCCATTTATTTT	TTCGTTGAGCAACTTAAAGTAGAGAATATTAGAGCGAA	3994						
Db 239457	AACCAAAATTTCTACAAAGATATTCACATTAATGATATAATATGT	TATAAGCATATC	239516						
QY 3995	AACCAAAATTTCTTCAAGATATTAACCTTATTTGATATAATAGATGTTAATAAGCATATC	4054							
Db 239517	TTGAATGAAAAAAATCAAAA	-ATCTGCGAAAC	239548						
QY 4055	TTGAATGAAAGTCAGCAAAATATGTGCGAAAC	4087							

RESULT	9				
LOCUS	CEF52G2	39271 bp	DNA	INV	18-DEC-1998
DEFINITION	Caenorhabditis elegans	cosmid F52G2,	complete sequence.		
ACCESSION	Z82269				
NID	93217492				
VERSION	Z82269.1	GI:3217492			
KEYWORDS	HTG.				
SOURCE	Caenorhabditis elegans.				
ORGANISM	Caenorhabditis elegans				
	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditodea; Rhabditidae; Peloderinae; Caenorhabditis				
REFERENCE	1 (bases 1 to 39271)				
AUTHORS	McMurray, A.				
TITLE	Direct Submission				
JOURNAL	Submitted (11-NOV-1996) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu				
REFERENCE	2 (bases 1 to 39271)				
AUTHORS	Wilson, R., Alnscough, R., Anderson, K., Baynes, C., Berks, M.,				

TITLE	2.2 Mb of contiguous nucleotide sequence from chromosome III of <i>C. elegans</i>
JOURNAL	Nature 368 (6466), 32-38 (1994)
MEDLINE	94150718
COMMENT	On Jun 13, 1998 this sequence version replaced qi:2546905.

<http://webcage.sanger.ac.uk/cgi-bin/display?db=wormbase&class=Sequence&object=F52G2>
Current sequencing criteria for the *C. elegans* genome
sequencing consortium are that all bases are either sequenced

FEATURES

source

unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

This sequence is the entire insert of clone F52G2. The start of this sequence (1..104) overlaps with the end of sequence AL021488. The end of this sequence (39168..39271) overlaps with the start of sequence AL021487.

1. 39271
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="F52G2"
complement(2363. 4916)
/gene="F52G2.1b"
complement(2363. 5967)
/gene="F52G2.1a"
complement(join(2363. 2713,2762. 2938,4410. 4841,
5941. 5967))
/gene="F52G2.1a"
/note="similar to Bacterial mutT protein; cDNA EST
yk552h1.5 comes from this gene; cDNA EST EMBL:Cl2643
comes from this gene; cDNA EST yk281h9.5 comes from this
gene"

```

/protein_id="CAB05206.1"
/db_xref="PID:e1354463"
/db_xref="PID:g3947559"
/db_xref="GI:3947559"
/db_xref="SPTREMBL:O62257"
/translation="MASNSTNSKKONKSTEEPSSVQKLLASLQQAONKSDISEQST
SKRRKNEKKKAVAQAPASAPAPGPEEKKQPKRASVGARMQQAENARISOTKRPQ
VUSKSSRRNTTAPEDQNYQQQQQYKGPRIPTDILDEFRISNVECEINDINIRV
CFHLELAHWYIDHMEVDKISGCPNVGSRDENQMCQHCRLVRYAHRADVLAKFR
EKSTVPTYGAILVDPMDHVLLQSYFAKGNMGFPKGINQAEPADARARETFEE
TGDEGIIYSEKEKKFORFINDGMVRLYLKKNVPDENFQOPOTRKEIRFEFD"
complement(join(2363..2713,2762..2938,4410..4916))

```

/note="similar to Bacterial mutr protein; CDNA EST
 yk357h8.5 comes from this gene"
 /codon_start=1
 /protein_id="CAB05204.1"
 /db_xref="PID:e1354461"
 /db_xref="PID:g3924774"
 /db_xref="GI:3924774"
 /db_xref="SP|REMBL:O62255"
 /translation="MEISTENWCKKPKNRSIFSKNISFQKQNKSTEEPPSSVQKLLAS
 LQQAQNKSDLSEQPTSTKPKKNEKTKKAAVAQAPASAPAPGEEKKQPPRASVGARQ
 QOAEARISQTKRPROVSTSKGSSRRKTAPEQNOOQOQYKGPRIPIIDLLEFR
 FISMVECEINDNIRVCFHLELAHWYIIDHVEEDKISGCPNNGSRDFNEQMOCQCRV
 LRKYAHRADEVLAKEFEYKSTVPTYGALIVDEMDHVLVQSYFAKKNWGFPPKGIN
 CAEPPRDAAIRETFETGTGDFGIYSEKEKKTQRFINDGMVRLYLKVNPKDFNFQPT
 RKEIRFGFLD"
 complement(6222..23742)

CDS

```
complement(join(62222..6257,7000..7560,8484..8723,
9864..10040,10106..10208,10261..10416,13378..13486,
13537..13712,13763..13890,14848..14953,15007..15170,
15756..15972,16439..16717,16765..16941,18397..18762,
21479..21621,21677..21809,22667..22801,23304..23441,
23489..23742))
/gene="F52G2.2"
/note="CDNA EST yk480b8.5 comes from this gene; CDNA EST
EMBL:D64786 comes from this gene; CDNA EST EMBL:D67931
comes from this gene; CDNA EST EMBL:D67947 comes from this
gene; CDNA EST EMBL:C10835 comes from this gene; CDNA EST
EMBL:C12654 comes from this gene; CDNA EST yk204g11.3
comes from this gene; CDNA EST yk209a7.3 comes from this
gene; CDNA EST yk209a7.5 comes from this gene; CDNA EST
```


QY 3815 CGGTGATGTAAAAAATGTCGATATATGTA-AAAAATGCATGCGTTTTCACAC-C--T 3870
Db 45008 TTCTACACACATGATAAAGAAAGTGTATATAAATACATTTT-GTATTTTCGTCA 45066
QY 3871 TTCTGCACAAATGATAGGGGAAATGTATTAAATACATTTTGTATTTTCAACA 3930
Db 45067 TCACATGAATAACCCCAT 45084
QY 3931 TCACATGATTAACCCCAT 3948
RESULT 6
LOCUS CEY62E10 256941 bp DNA HTG 12-MAR-1999
DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
clone Y62E10, WORKING DRAFT SEQUENCE.
ACCESSION AL031580
NID 94469033
VERSION AL031580.4 GI:4469033
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS 1 (bases 1 to 256941)
Lloyd, C.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-1999) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwenematode.wustl.edu
COMMENT On Mar 22, 1999 this sequence version replaced gi:4455386.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, Yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
source location/Qualifiers
1..256941
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="Y62E10"
BASE COUNT 77397 a 44556 c 43039 g 79104 t 12845 others
ORIGIN
Query Match 15.4%; Score 154; DB 19; Length 256941;
Best Local Similarity 85.7%; Pred. No. 4.67e-79;
Matches 221; Conservative 0; Mismatches 31; Indels 6; Gaps 5;
Db 113063 CATTTTCTACATTTTCTACATGACCGACAGTCCACTTCTCATGCTGAAGACGAGAGA 113122
QY 3696 CATTTTTCGACATTTT-TACATCGAAGACGACAGCTCACTTCACATGCTGAAGACGAGAGA 3754
Db 113123 CGCAGAGAAATACACACATCTCTGCGTCTCTCGCCTTCAGCATTTAAAGTGAGCTGT 113182
QY 3755 CGCGGAGAAATACACACATCTCTGCGTCTCTCGTCTTCAGCATGTGAATGGGATCT 3814
Db 113183 CGGCCGATGTAGAAAAATTCGAAAAATGTAGAAAAATGCATGCGTTTAAAAACATTT 113242
QY 3815 CGGTCCGATGTAAAAAATGTCGATTAATGTA-AAAAATGCATGCGTTTTCACA-C--T 3870
Db 113243 TTCTACACACATGATAAAGAAAGTGTATATAAATACATTTT-GTATTTTTCGTCA 113301
QY 3871 TTCTGCACAAATGATAGGGGAAATGTATTAAATACATTTTGTATTTTTCACA 3930
Db 113302 TCACATGAATAACCCCAT 113319

QY 3931 TCACATGATTAACCCCAT 3948
RESULT 7
LOCUS CEC48D1 39908 bp DNA INV 23-NOV-1998
DEFINITION Caenorhabditis elegans cosmid C48D1, complete sequence.
ACCESSION Z81049
NID 91627677
VERSION Z81049.1 GI:1627677
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS 1 (bases 1 to 39908)
Burton, J.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwenematode.wustl.edu
2 (bases 1 to 39908)
REFERENCE Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,
Saunders, D., Showken, R., Smaiden, N., Smith, A., Sonhammer, E.,
Vaughan, K., Waterston, R., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Wilkinson-Sproat, J., and Wohlman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
COMMENT Coding sequences below are predicted from computer analysis, using
predictions from Genefinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:-
http://webace.sanger.ac.uk/cgi-
bin/display?db=wormace&class=Sequence &object=C48D1
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we arrange for a small
overlapping sections once, or longer because we only sequence
overlap between neighbouring submissions.
This sequence is the entire insert of clone C48D1. The true right
end of clone F58D2 is at 18510 in this sequence. The start of this
sequence (1..100) overlaps with the end of sequence Z81093.
The end of this sequence (33020..39908) overlaps with the start of
sequence Z82274.
FEATURES
source location/Qualifiers
1..39908
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="C48D1"
complement(94..4439)
/gene="C48D1.1"
complement(join(94..1020,1345..1951,2283..2580,
2629..2765,2882..3147,3615..3739,4089..4331,4382..4439))
/gene="C48D1.1"
/note="similar to C.elegans ZYG-11 like protein; cDNA EST
EMBL:Z14909 comes from this gene; cDNA EST EMBL:T01768
comes from this gene"
/codon_start=1
/protein_id="CAB02846.1"


```

/db_xref="PID:93924776"
/db_xref="GI:3924776"
/translation="MTEDGVMKIDASTSTIKEPPADCVILPHTNEOVVYKSMIAF
GDNPNLYTNRLDGLADWYIVDSDSPPTGPVYKAFIYVNDLNOESGNPLFKVTG
RLKFLANKDTSDFMKLEIFEONRLLELRSPKMSDSQVIAIILEDNEETNSAICYS
EOLESSIKVHLPGVSPAKAGQILIFTEYNYENDTYEPDRPDALIEYRECEMTRIEAN
LISAKAACSSIEHTITYPOSAFADTFGKVNAAESMERGVTRYVELAINIIDSEAN
NVLPLKISDIGEKLTEGENEALFLKNDLFEELQWEDMTPEELAVPEPVAAEPA
DPEVEAKIEENGVNDESLQESDVLFLQIESASSEETMESNSAENSEKPKTFE
GMPQSLSESSSSSTISSKGVDSVIVLVNDECVGFETQKLMTICTAGVTPL
RYGIEFOEIAAGNNGKTSIQOSARVSECEGAVFSSDDPKISANVITSSNPKHKS
NKEVGVSDRYGVEIPRTGDEFEGTVSTQITWYQGDPLKACFKAVAPLKRQFIDD
LDHYLVGMKYEDNIYOLSRKNGLGAKSISRLPNEPRNGSKHSGTKRFTTPHTGLV
DMERKGNYPGCVWTCFGDVYFKSTPTFSQNGYPSVGDVNIIVEILSPGDKNLG
ASMTKEPASLSTVIVRNNGEIVKAYLIMKPEKLIYSEDNNRYEHDLQIVVADSSI
YGSDEQSKVFEERLNPITLIGFTTNLMKCYMAARLPMIHREPSPEPIKNDEN
VYDPEQKGPQSVNSYRSDRQYQEKPMOYQSYHDPADTISISQDNHREMETRNSAG
TSVSSVDRKEVGVVMKVDKNTQIVYLSGQKAAIHHWNTHSOSTIVGNSYIEMVPI
EVHKAHYAPEITQVVEILNTEHGRVDRSEFSLFTDLNVDQRYVERLAFGEAL
VLRSDVGHVVMASNVKTPLNMTAVEVCQTYKGRSENLVGYCRVYGVKPREIV
DGHFSDQFVSAYYFELHEILATSNAYIVRNENLDRNLRGNTIEVRDDISQKRIAH
MNFGEQOQOQSSLSRGPTRYSPVGSYQSSHQNDNRSAAHNGDGRSTOSHSAA
GGSNMHONFYQPPVYQSSSQOSQOPRYNGAPSGPLPPSSQSSFGAPAPAPHSEMS
NDYNRMVYKLAELDTAAQMIINNDIRQIKNISPLDLDERNIDACQREKRYEYK
YGN"
complement(27575..32349)
/gene="F52G2.3"
complement(join(27575..27979,29638..29814,29857..29959,
30010..30132,30177..30380,30607..30741,31300..31446,
31498..31803,32240..32349))
/gene="F52G2.3"
/codon_start=1
/protein_id="CAB05205.1"
/db_xref="PID:e1354462"
/db_xref="PID:93924775"
/db_xref="GI:3924775"
/db_xref="SPTREMBL:062256"
/translation="METRLVLKYDILNREAICLNRETHELHANLVESVOPQIGEL
AVENLKIIVDGVCFISPEGTVIEPPAPEDCVILPEKEDVYVMKSWIGFSSDSPH
LTYDRNIGYADWYGPVNCOSQFKAKNTKYAFKYLIAVNDLDYKNTGECIMKAVQLVGS
IDNRDAEVRLEKMEFERKRIADIQRTKQRLIETVAYPIPSAHQSPSSNHSYTGRS
SCDSLRPQVPAPARKLSEOPQIVONQAPYREQINOQSVTNSIYGEFSGVGRDQPLE
TANSRGSTEYSGRSSSSSDQVIRNEPQLPNPAPILPESMGWYTYENGIMQIVYLRV
QKAIHYWSKHQITSLGVSFNCSEFKPIPVQATTHWAPETAITANNVNEKNEKIVG
KECFALFKVDLSRPEGKFSKWTITIGNPJFSCIEIVYMLAHLSDTADNVKRHNDI
EMOKKRAQOEHPAQOORPAGSSNSYQTENAOQHSMSNMQYHSGAPPVSSRSSISLG
ETAEFKNKLERTNDNMESLCAIKKFLDNRLQKETLRGYNPRRLQALEEKLKESHL
ENN"
complement(33889..34904)
/gene="F52G2.4"
complement(join(33889..33909,33958..34013,34063..34141,
34714..34802,34853..34904))
/gene="F52G2.4"
/translation="CDNA EST yk356f5.5 comes from this gene; CDNA EST
EMBL:D64203 comes from this gene; CDNA EST EMBL:D66267
comes from this gene; CDNA EST yk298f3.3 comes from this
gene; CDNA EST yk356f5.3 comes from this gene; CDNA EST
yk383e11.3 comes from this gene"
/codon_start=1
/protein_id="CAB05208.1"
/db_xref="PID:e1354465"
/db_xref="PID:93924777"
/db_xref="GI:3924777"
/translation="MRGVEENDGLVMTMCRVVERKKDENMKWRDESGRLDEKHHN
TWKGEIDTLRLDGPNEGHRHSTVEKMPGTDRFDVNRQYNKEYVESIYNY"
BASE COUNT 12956 a 6666 c 6504 g 13145 t
ORIGIN
Query Match 15.4%; Score 154; DB 21; Length 39271;
Best Local Similarity 85.7%; Pred. No. 4.67e-79;
Matches 221; Conservative 0; Mismatches 31; Indels 6; Gaps 5;
Db 3227 CATTTTCTACATTTTCTACATCGACCGACAGTCACCTTCTCATGCTGAAGACGAGAGA 3286

/db_xref="PID:93924776"
/db_xref="GI:3924776"
/translation="MTEDGVMKIDASTSTIKEPPADCVILPHTNEOVVYKSMIAF
GDNPNLYTNRLDGLADWYIVDSDSPPTGPVYKAFIYVNDLNOESGNPLFKVTG
RLKFLANKDTSDFMKLEIFEONRLLELRSPKMSDSQVIAIILEDNEETNSAICYS
EOLESSIKVHLPGVSPAKAGQILIFTEYNYENDTYEPDRPDALIEYRECEMTRIEAN
LISAKAACSSIEHTITYPOSAFADTFGKVNAAESMERGVTRYVELAINIIDSEAN
NVLPLKISDIGEKLTEGENEALFLKNDLFEELQWEDMTPEELAVPEPVAAEPA
DPEVEAKIEENGVNDESLQESDVLFLQIESASSEETMESNSAENSEKPKTFE
GMPQSLSESSSSSTISSKGVDSVIVLVNDECVGFETQKLMTICTAGVTPL
RYGIEFOEIAAGNNGKTSIQOSARVSECEGAVFSSDDPKISANVITSSNPKHKS
NKEVGVSDRYGVEIPRTGDEFEGTVSTQITWYQGDPLKACFKAVAPLKRQFIDD
LDHYLVGMKYEDNIYOLSRKNGLGAKSISRLPNEPRNGSKHSGTKRFTTPHTGLV
DMERKGNYPGCVWTCFGDVYFKSTPTFSQNGYPSVGDVNIIVEILSPGDKNLG
ASMTKEPASLSTVIVRNNGEIVKAYLIMKPEKLIYSEDNNRYEHDLQIVVADSSI
YGSDEQSKVFEERLNPITLIGFTTNLMKCYMAARLPMIHREPSPEPIKNDEN
VYDPEQKGPQSVNSYRSDRQYQEKPMOYQSYHDPADTISISQDNHREMETRNSAG
TSVSSVDRKEVGVVMKVDKNTQIVYLSGQKAAIHHWNTHSOSTIVGNSYIEMVPI
EVHKAHYAPEITQVVEILNTEHGRVDRSEFSLFTDLNVDQRYVERLAFGEAL
VLRSDVGHVVMASNVKTPLNMTAVEVCQTYKGRSENLVGYCRVYGVKPREIV
DGHFSDQFVSAYYFELHEILATSNAYIVRNENLDRNLRGNTIEVRDDISQKRIAH
MNFGEQOQOQSSLSRGPTRYSPVGSYQSSHQNDNRSAAHNGDGRSTOSHSAA
GGSNMHONFYQPPVYQSSSQOSQOPRYNGAPSGPLPPSSQSSFGAPAPAPHSEMS
NDYNRMVYKLAELDTAAQMIINNDIRQIKNISPLDLDERNIDACQREKRYEYK
YGN"
complement(27575..32349)
/gene="F52G2.3"
complement(join(27575..27979,29638..29814,29857..29959,
30010..30132,30177..30380,30607..30741,31300..31446,
31498..31803,32240..32349))
/gene="F52G2.3"
/codon_start=1
/protein_id="CAB05205.1"
/db_xref="PID:e1354462"
/db_xref="PID:93924775"
/db_xref="GI:3924775"
/db_xref="SPTREMBL:062256"
/translation="METRLVLKYDILNREAICLNRETHELHANLVESVOPQIGEL
AVENLKIIVDGVCFISPEGTVIEPPAPEDCVILPEKEDVYVMKSWIGFSSDSPH
LTYDRNIGYADWYGPVNCOSQFKAKNTKYAFKYLIAVNDLDYKNTGECIMKAVQLVGS
IDNRDAEVRLEKMEFERKRIADIQRTKQRLIETVAYPIPSAHQSPSSNHSYTGRS
SCDSLRPQVPAPARKLSEOPQIVONQAPYREQINOQSVTNSIYGEFSGVGRDQPLE
TANSRGSTEYSGRSSSSSDQVIRNEPQLPNPAPILPESMGWYTYENGIMQIVYLRV
QKAIHYWSKHQITSLGVSFNCSEFKPIPVQATTHWAPETAITANNVNEKNEKIVG
KECFALFKVDLSRPEGKFSKWTITIGNPJFSCIEIVYMLAHLSDTADNVKRHNDI
EMOKKRAQOEHPAQOORPAGSSNSYQTENAOQHSMSNMQYHSGAPPVSSRSSISLG
ETAEFKNKLERTNDNMESLCAIKKFLDNRLQKETLRGYNPRRLQALEEKLKESHL
ENN"
complement(33889..34904)
/gene="F52G2.4"
complement(join(33889..33909,33958..34013,34063..34141,
34714..34802,34853..34904))
/gene="F52G2.4"
/translation="CDNA EST yk356f5.5 comes from this gene; CDNA EST
EMBL:D64203 comes from this gene; CDNA EST EMBL:D66267
comes from this gene; CDNA EST yk298f3.3 comes from this
gene; CDNA EST yk356f5.3 comes from this gene; CDNA EST
yk383e11.3 comes from this gene"
/codon_start=1
/protein_id="CAB05208.1"
/db_xref="PID:e1354465"
/db_xref="PID:93924777"
/db_xref="GI:3924777"
/translation="MRGVEENDGLVMTMCRVVERKKDENMKWRDESGRLDEKHHN
TWKGEIDTLRLDGPNEGHRHSTVEKMPGTDRFDVNRQYNKEYVESIYNY"
BASE COUNT 12956 a 6666 c 6504 g 13145 t
ORIGIN
Query Match 15.4%; Score 154; DB 21; Length 39271;
Best Local Similarity 85.7%; Pred. No. 4.67e-79;
Matches 221; Conservative 0; Mismatches 31; Indels 6; Gaps 5;
Db 3227 CATTTTCTACATTTTCTACATCGACCGACAGTCACCTTCTCATGCTGAAGACGAGAGA 3286

RESULT 5
LOCUS CEY45F10 434238 bp DNA HTG 03-DEC-1998
DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
clone Y45F10, WORKING DRAFT SEQUENCE.
ACCESSION Z93245
NID 93378089
VERSION Z93245.1 GI:3378089
KEYWORDS HTG, HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 434238)
AUTHORS McMurray, A.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1998) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanjour.ac.uk or twenematode.wustl.edu
On Aug 4, 1998 this sequence version replaced gi:1906319.
Order of segments is not known; 800 n's separate segments.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
Source Location/Qualifiers
1..434238
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="Y45F10"
BASE COUNT 141590 a 73465 c 71957 g 143994 t 3232 others
ORIGIN
Query Match 15.4%; Score 154; DB 19; Length 110000;
Best Local Similarity 85.7%; Pred. No. 4.67e-79;
Matches 221; Conservative 0; Mismatches 31; Indels 6; Gaps 5;
Db 44828 CATTTTCTACATTTTCTACATCGACCGACAGTCACCTTCTCATGCTGAAGACGAGAGA 44887
QY 3696 CATTTTTCGACATTTT-TACATCGAACGACAGCTCACCCTCACATGCTGAAGACGAGAGA 3754
Db 44888 CGCAGAGAAATACACACATCTCTCGCTCTCTCTCGCTTACGACATTTAAAGTGAGCTGT 44947
QY 3755 CGCGAGAGAAATACACACATCTTCTGCGTCTCTCTGCTTTCAGCATGTGAATGGGATCT 3814
Db 44888 CGCAGAGAAATACACACATCTCTCGCTCTCTCTCGCTTACGACATTTAAAGTGAGCTGT 44947
QY 3755 CGCGAGAGAAATACACACATCTTCTGCGTCTCTCTGCTTTCAGCATGTGAATGGGATCT 3814
Db 44948 CGGCCGATGTAGAAAAATTTGAAAAATGTAGAAAAATGCATGCGTTTAAAAACATTT 45007

```



```

/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="F52G2"
complement(2363. .4916)
/gene="F52G2.1b"
complement(2363. .5967)
/gene="F52G2.1a"
complement(join(2363. .2713,2762. .2938,4410. .4841,
5941. .5967))
/gene="F52G2.1a"
/note="similar to Bacterial mutT protein; cDNA EST
yk52h11.5 comes from this gene; cDNA EST EMBL:C12643
comes from this gene; cDNA EST yk281h9.5 comes from this
gene"

```

```

/db_xref="PID:e1354463"
/db_xref="PID:g3947559"
/db_xref="GI:3947559"
/db_xref="SPTREMBL:O62257"
/translation="MASNSTNSKONKSTEEPPSSVOKLLASLQQAQNKSDLSEQBST
SKPKNEKRKRAVAQAAPASAPAPGPEEKKKQPKRASVGAQMQQQAENARISQTKPRQ
VSKSKSSRNTPAPEQOONYQQQOQOQKGPRIPTDILDELEFISNMVECEINDNIRV
CEHLELAHWYIDHMAVEDDKISGCNPNVSRDENFQMCQHCRLVKRYAHRADLEYLAKER
EYKSTVPTYGAILVDPEMDHVVLYIOSYFAKGNMGFKPKGINOAQPPRDAITPEEER

```

```

/gene="F52G2.1b"
/note="similar to Bacterial mult protein; CDNA EST
YK357h8.5 comes from this gene"
/codon_start=1
/protein_id="CAB05204.1"

```

gene

complement(6222. .23742)

/gene="F52G2.2"

CDS

complement(join(6222. .6257,7000. .7560,8484. .8723,
9864. .10040,10106. .10208,10261. .10416,13378. .13486,
13537. .13712,13763. .13890,14848. .14953,15007. .15170,
15156. .15972,16439. .16717,16765. .16941,18397. .18762,
18762. .18762,18763. .18763,18764. .18764,18765. .18765,18766. .18766,18767. .18767,18768. .18768,18769. .18769,18770. .18770,18771. .18771,18772. .18772,18773. .18773,18774. .18774,18775. .18775,18776. .18776,18777. .18777,18778. .18778,18779. .18779,18780. .18780,18781. .18781,18782. .18782,18783. .18783,18784. .18784,18785. .18785,18786. .18786,18787. .18787,18788. .18788,18789. .18789,18790. .18790,18791. .18791,18792. .18792,18793. .18793,18794. .18794,18795. .18795,18796. .18796,18797. .18797,18798. .18798,18799. .18799,18800. .18800,18801. .18801,18802. .18802,18803. .18803,18804. .18804,18805. .18805,18806. .18806,18807. .18807,18808. .18808,18809. .18809,18810. .18810,18811. .18811,18812. .18812,18813. .18813,18814. .18814,18815. .18815,18816. .18816,18817. .18817,18818. .18818,18819. .18819,18820. .18820,18821. .18821,18822. .18822,18823. .18823,18824. .18824,18825. .18825,18826. .18826,18827. .18827,18828. .18828,18829. .18829,18830. .18830,18831. .18831,18832. .18832,18833. .18833,18834. .18834,18835. .18835,18836. .18836,18837. .18837,18838. .18838,18839. .18839,18840. .18840,18841. .18841,18842. .18842,18843. .18843,18844. .18844,18845. .18845,18846. .18846,18847. .18847,18848. .18848,18849. .18849,18850. .18850,18851. .18851,18852. .18852,18853. .18853,18854. .18854,18855. .18855,18856. .18856,18857. .18857,18858. .18858,18859. .18859,18860. .18860,18861. .18861,18862. .18862,18863. .18863,18864. .18864,18865. .18865,18866. .18866,18867. .18867,18868. .18868,18869. .18869,18870. .18870,18871. .18871,18872. .18872,18873. .18873,18874. .18874,18875. .18875,18876. .18876,18877. .18877,18878. .18878,18879. .18879,18880. .18880,18881. .18881,18882. .18882,18883. .18883,18884. .18884,18885. .18885,18886. .18886,18887. .18887,18888. .18888,18889. .18889,18890. .18890,18891. .18891,18892. .18892,18893. .18893,18894. .18894,18895. .18895,18896. .18896,18897. .18897,18898. .18898,18899. .18899,18900. .18900,18901. .18901,18902. .18902,18903. .18903,18904. .18904,18905. .18905,18906. .18906,18907. .18907,18908. .18908,18909. .18909,18910. .18910,18911. .18911,18912. .18912,18913. .18913,18914. .18914,18915. .18915,18916. .18916,18917. .18917,18918. .18918,18919. .18919,18920. .18920,18921. .18921,18922. .18922,18923. .18923,18924. .18924,18925. .18925,18926. .18926,18927. .18927,18928. .18928,18929. .18929,18930. .18930,18931. .18931,18932. .18932,18933. .18933,18934. .18934,18935. .18935,18936. .18936,18937. .18937,18938. .18938,18939. .18939,18940. .18940,18941. .18941,18942. .18942,18943. .18943,18944. .18944,18945. .18945,18946. .18946,18947. .18947,18948. .18948,18949. .18949,18950. .18950,18951. .18951,18952. .18952,18953. .18953,18954. .18954,18955. .18955,18956. .18956,18957. .18957,18958. .18958,18959. .18959,18960. .18960,18961. .18961,18962. .18962,18963. .18963,18964. .18964,18965. .18965,18966. .18966,18967. .18967,18968. .18968,18969. .18969,18970. .18970,18971. .18971,18972. .18972,18973. .18973,18974. .18974,18975. .18975,18976. .18976,18977. .18977,18978. .18978,18979. .18979,18980. .18980,18981. .18981,18982. .18982,18983. .18983,18984. .18984,18985. .18985,18986. .18986,18987. .18987,18988. .18988,18989. .18989,18990. .18990,18991. .18991,18992. .18992,18993. .18993,18994. .18994,18995. .18995,18996. .18996,18997. .18997,18998. .18998,18999. .18999,19000. .19000,19001. .19001,19002. .19002,19003. .19003,19004. .19004,19005. .19005,19006. .19006,19007. .19007,19008. .19008,19009. .19009,19010. .19010,19011. .19011,19012. .19012,19013. .19013,19014. .19014,19015. .19015,19016. .19016,19017. .19017,19018. .19018,19019. .19019,19020. .19020,19021. .19021,19022. .19022,19023. .19023,19024. .19024,19025. .19025,19026. .19026,19027. .19027,19028. .19028,19029. .19029,19030. .19030,19031. .19031,19032. .19032,19033. .19033,19034. .19034,19035. .19035,19036. .19036,19037. .19037,19038. .19038,19039. .19039,19040. .19040,19041. .19041,19042. .19042,19043. .19043,19044. .19044,19045. .19045,19046. .19046,19047. .19047,19048. .19048,19049. .19049,19050. .19050,19051. .19051,19052. .19052,19053. .19053,19054. .19054,19055. .19055,19056. .19056,19057. .19057,1

```

23403. 1231421)
/gene="F52G2.2"
/note="CDNA EST yk480b8.5 comes from this gene; CDNA EST
EMBL:D64786 comes from this gene; CDNA EST EMBL:D67931
comes from this gene; CDNA EST EMBL:D67947 comes from this
gene; CDNA EST EMBL:C10835 comes from this gene; CDNA EST
EMBL:C12654 comes from this gene; CDNA EST yk204g11.3
comes from this gene; CDNA EST yk209a7.3 comes from this
gene; CDNA EST yk209a7.5 comes from this gene; CDNA EST
yk210a2.3 comes from this gene; CDNA EST yk216a4.3 comes
from this gene; CDNA EST yk216a4.5 comes from this gene;
CDNA EST yk226f6.5 comes from this gene; CDNA EST
yk338a3.3 comes from this gene; CDNA EST yk338a3.5 comes
from this gene; CDNA EST yk351f3.5 comes from this gene;
CDNA EST yk406e11.5 comes from this gene; CDNA EST
yk408e8.3 comes from this gene; CDNA EST yk408e8.5 comes
from this gene; CDNA EST yk427f8.5 comes from this gene;
CDNA EST yk480b8.3 comes from this gene; CDNA EST
EMBL:T00333 comes from this gene"

```

Location/Qualifiers
1. .39271

Cp 4360 TTACGACTTGAACTGTATCCGAGAGAGAGACATCCCGTGAATGAAGAGTTGGCGATGCT 4301
Db 9062 GAAATGTGTGTTTATAGATCCGATTTTGATACCCAAATTTCCCTTTTGGACGTTTCG 9121
Cp 4300 GAAATGTGTGTTTATAGATCCGATTTTGATACCCAAATTTCCCTTTTGGACGTTTCG 4241
Db 9122 CACATTTTAATTTGCGCAATGTTTCATTCATTAATATCTCTAGAAATTTTGATTTCCCTT 9181
Cp 4240 CACATTTTAATTTGCGCAATGTTTCATTCATTAATATCTCTAGAAATTTTGATTTCCCTT 4181
Db 9182 TTAATTTCTACATTCATGATGTAGAAAAATGCAAAATGTATTTTAATGCAATTTT 9241
Cp 4180 TTAATTTCTACATTCATGATGTAGAAAAATGCAAAATGTATTTTAATGCAATTTT 4121
Db 9242 CAATTTTCGAGAAATTTTGAATTTTTCAGGTGTTTCGACATATTTTGGTGAATTTTC 9301
Cp 4120 CAATTTTCGAGAAATTTTGAATTTTTCAGGTGTTTCGACATATTTTGGTGAATTTTC 4061
Db 9302 ATTCAGATATGCTTATTAACATCTATATTAATCAATAAGTAATATCTTGAAGAAAT 9361
Cp 4060 ATTCAGATATGCTTATTAACATCTATATTAATCAATAAGTAATATCTTGAAGAAAT 4001
Db 9362 TTGTTTTCGCTCTAATATTTCTCTACTTTTAAAGTGTCAACGAAAAATATGGGTT 9421
Cp 4000 TTGTTTTCGCTCTAATATTTCTCTACTTTTAAAGTGTCAACGAAAAATATGGGTT 3941
Db 9422 AATCATGTGATGTGAAAAATACAAAAATGTATTTTAATACATTTTCCCTATTCAT 9481
Cp 3940 AATCATGTGATGTGAAAAATACAAAAATGTATTTTAATACATTTTCCCTATTCAT 3881
Db 9482 TGTGCAGAAAAAGTGTAAAAAACGCGATGCTTTTACATTAATGACATTTTTCAT 9541
Cp 3880 TGTGCAGAAAAAGTGTAAAAAACGCGATGCTTTTACATTAATGACATTTTTCAT 3821
Db 9542 CGACCGACATCCCATTTTCACATGCTGAAGAGAGAGACGAGAAAGATGTGTGATTTTC 9601
Cp 3820 CGACCGACATCCCATTTTCACATGCTGAAGAGAGAGACGAGAAAGATGTGTGATTTTC 3761
Db 9602 TCCGCGCTCTCTGCTTCAGCATGTGAAGTGAAGTGTGCTTGCATGTAAAAAATGTCGAA 9661
Cp 3760 TCCGCGCTCTCTGCTTCAGCATGTGAAGTGAAGTGTGCTTGCATGTAAAAAATGTCGAA 3701
Db 9662 AAATGCATGCGCTTTTAC-ATTTTCTGCACA-CATGAATAGGGGAAAAATATATAAA 9719
Cp 3700 AAATGCATGCGCTTTTAC-ATTTTCTGCACA-CATGAATAGGGGAAAAATATATAAA 3641
Db 9720 AAT-GC-ATTTTCTGCAT-CATATGAACAA-CCACATGA 9755
Cp 3640 AATGCTATTCTTCTCATTCATATGAACAAACACATGA 3601

RESULT 3
LOCUS CEV67H2 314495 bp DNA HTG 04-MAR-1999
DEFINITION Caenorhabditis elegans DNA *** SEQUENCING IN PROGRESS *** from
clone Y67H2, WORKING DRAFT SEQUENCE.
ACCESSION AL022475
NID 94469034
VERSION AL022475.3 GI:4469034
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 314495)
AUTHORS McMurtry, A.
TITLE Direct Submission
JOURNAL Submitted (04-MAR-1999) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RO, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@anger.ac.uk or rwenematode.wustl.edu

COMMENT On Mar 22, 1999 this sequence version replaced gi:4468145.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress

and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES

source

1..314495
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="Y67H2"

BASE COUNT 96542 a 54551 c 52760 g 97027 t 13615 others
ORIGIN

Query Match 93.8%; Score 938; DB 19; Length 314495;
Best Local Similarity 98.4%; Pred. No. 0.00e+00;
Matches 984; Conservative 0; Mismatches 10; Indels 6; Gaps 6;

Db 238419 TAGCCCATGATCTGAACAAATTTGTAAAGATTGCTGTCGCTTGTGATCCATCCGT 238478
Cp 4600 TAGCCCATGATCTGAACAAATTTGTAAAGATTGCTGTCGCTTGTGATCCATCCGT 4541
Db 238479 GTTGCATCTGCTCAAAAGTGTTCATTAATTTATGATGAGGCACATTCACGAGACTCGAG 238538
Cp 4540 GTTGCATCTGCTCAAAAGTGTTCATTAATTTATGATGAGGCACATTCACGAGACTCGAG 4481
Db 238539 AAGTTCTGTACATGTTTCTCTGTCGCAAAACACGGCTTATGTTGGTGCATCGACAAAG 238598
Cp 4480 AAGTTCTGTACATGTTTCTCTGTCGCAAAACACGGCTTATGTTGGTGCATCGACAAAG 4421
Db 238599 TTCTATCTCTTCATGGAATATGTATGAGTTGTTCCAGAAAGCTTTGCTGATGAGCGA 238658
Cp 4420 TTCTATCTCTTCATGGAATATGTATGAGTTGTTCCAGAAAGCTTTGCTGATGAGCGA 4361
Db 238659 TTACGACTTGAACCTGATCCGAGAGAGAGATCCGGTGAATGAAGACTTGGCGGATGCT 238718
Cp 4360 TTACGACTTGAACCTGATCCGAGAGAGAGATCCGGTGAATGAAGACTTGGCGGATGCT 4301
Db 238719 GAAATGTGTGTTTAAAGATCGATTTTGATACCCAAATTTCCCTCTTTTGAAGCTTTCG 238778
Cp 4300 GAAATGTGTGTTTAAAGATCGATTTTGATACCCAAATTTCCCTCTTTTGAAGCTTTCG 4241
Db 238779 CACATTTTAATTTGCGAATGTTTCATTCATTAATCTCTAGAAATTTTGAATTTCCCTT 238838
Cp 4240 CACATTTTAATTTGCGAATGTTTCATTCATTAATCTCTAGAAATTTTGAATTTCCCTT 4181
Db 238839 TTAATTTTCTACATTCATGATGTAGAAAAATGCAAAATGTACGGTAAACCATTTGTT 238898
Cp 4180 TTAATTTTCTACATTCATGATGTAGAAAAATGCAAAATGTATTTTAATGCAATTTT 4121
Db 238898 CAATTTTCGAGAAATTTTGAATTTTTCAGGTGTTTCGACATATTTTGGTGAATTTTC 238958
Cp 4120 CAATTTTCGAGAAATTTTGAATTTTTCAGGTGTTTCGACATATTTTGGTGAATTTTC 4061
Db 238959 ATTCAGATATCCGATTAATCTATAATTAATCAATAAGCAATATCTTGAAGAAAT 239018
Cp 4060 ATTCAGATATCCGATTAATCTATAATTAATCAATAAGCAATATCTTGAAGAAAT 4001
Db 239019 TTGTTTTCGCTCTAATATTTCTCTACTTTTAAAGTGTCAACGAAAAAATATGGGCTT 239078
Cp 4000 TTGTTTTCGCTCTAATATTTCTCTACTTTTAAAGTGTCAACGAAAAAATATGGGCTT 3941
Db 239079 AATCATGTGATGTGAAAAATACAAAAATGTATTTTAATACATTTTCCCTATTCAT 239138
Cp 3940 AATCATGTGATGTGAAAAATACAAAAATGTATTTTAATACATTTTCCCTATTCAT 3881
Db 239139 TGTGCAGAAAAAGTGTAAAAAACGCGATGCTTTTACATTAATTCGACATTTTTCAT 239198
Cp 3880 TGTGCAGAAAAAGTGTAAAAAACGCGATGCTTTTACATTAATTCGACATTTTTCAT 3821

FEATURES
source

Query Match	Best Local Similarity	Score	DB	Length
Matches 993; Conservative 0; Mismatches 1; Indels 6; Gaps 6;	95.6%;	956;	DB 21;	Length 39908;
Db 8762	TAGCCCATGATCGAACAATTGTAAGATTGCTTGTGCGGCTTGATCCATTCCT	8821		
Cp 4600	TAGCCCATGATCGAACAATTGTAAGATTGCTTGTGCGGCTTGATCCATTCCT	4541		
Db 8822	GTTGGCATCTGCTCAAAAGTGTTCATTTATGATGAGGCACATTCACGAGGACTCGAG	8881		
Cp 4540	GTTGGCATCTGCTCAAAAGTGTTCATTTATGATGAGGCACATTCACGAGGACTCGAG	4481		
Db 8882	AAGTTCTGTACATGTTTCTCGTCGAAAAACACGGCTTATGTTGGTGCATCGACAAAG	8941		
Cp 4480	AAGTTCTGTACATGTTTCTCGTCGAAAAACACGGCTTATGTTGGTGCATCGACAAAG	4421		
Db 8942	TTCATATCTCTTCATGGAATATGATTTGAGTTGGTCCAGAACTTTGCTGAATGAGCGA	9001		
Cp 4420	TTCATATCTCTTCATGGAATATGATTTGAGTTGGTCCAGAACTTTGCTGAATGAGCGA	4361		
Db 9002	TTACGACTTGAAGTGTATCCGAGAGAGCATCCGGTGAATGAAGAGTTGGCGGATGCT	9061		

/protein_id="AAA27982.1"
/db_xref="PID:9456417"
/db_xref="GI:456417"
/translation="MMDRDRSLERINIMFSSHLKVEILEVLIAKOVLSNDGDMINSCGTREKREIVKAVQRQDVAFDAFYDALRSTGHEGLAEVLEPLARSVDNAVEECPMPASHRSRALSPAGYSPTRVHRDSVSSSFTSYODIYSRARSRSRALHS SDRHNSPPVNAFSPQSSANSSTGCSLGYSSSRNRSFKASGPTQYTFHEEDMN FVDAPISRVEDEKTMYNFSSPRGMCLINNEHEFQMPTRNGTKADKDNLNLFCM GYTVICKDNLTRGMLTIRDFAKHESHGDSAILVLSHGEENVIGVDDIPISTHEI YDLNANAPRLANKPKIVFQACRGERRDNGFPVLDSDVGPAPFLRGWDRDPLF NFLGCVRPQOVQWRKPKPSQADILIRYATTAQYVSWRNSRSGSWFIQAVCEVFSTHAK DMVVELLITEVNKKVACGFQTSQGSNILKOMPEMSTRLLKFFFWPEARNSAV"
2367. : 2429
/gene="ced-3"
/number=1
exon 2430. : 2575
/gene="ced-3"
/number=2
intron 2576. : 2853
/gene="ced-3"
/number=2
exon 2854. : 3107
/gene="ced-3"
/number=3
intron 3108. : 4302
/gene="ced-3"
/number=3
exon 4303. : 4634
/gene="ced-3"
/number=4
intron 4635. : 5546
/gene="ced-3"
/number=4
exon 5547. : 5760
/gene="ced-3"
/number=5
intron 5761. : 5814
/gene="ced-3"
/number=5
exon 5815. : 5942
/gene="ced-3"
/number=6
intron 5943. : 6297
/gene="ced-3"
/number=6
exon 6298. : 6537
/gene="ced-3"
/number=7
intron 6538. : 7012
/gene="ced-3"
/number=7
exon 7013. : 7652
/gene="ced-3"
/number=8
BASE COUNT 2429 a 1455 c 1271 g 2498 t
ORIGIN

Query Match 100.0%; Score 1000; DB 21; Length 7653;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 3601 TCATGTTGTTGTTTCATATGAATGAGAAAAATAGCAATTTTATATATTTCCCTAT 3660
|||
QY 3601 TCATGTTGTTGTTTCATATGAATGAGAAAAATAGCAATTTTATATATTTCCCTAT 3660

DB 3661 TCATGTTGTCAGAAAAATAGTAAAAAGCGCATGATTTTTCGACATTTTTCATCGA 3720
|||
QY 3661 TCATGTTGTCAGAAAAATAGTAAAAAGCGCATGATTTTTCGACATTTTTCATCGA 3720

DB 3721 ACGACAGCTCACTTCATGCTGAAGACGAGAGACGCGGAGAAATACACACATCTTCT 3780
|||
QY 3721 ACGACAGCTCACTTCATGCTGAAGACGAGAGACGCGGAGAAATACACACATCTTCT 3780

DB 3781 GCGTCTCTGCTCTTCAGCATGTGAATGGGATCTCGGTCGATGTAAAAAATGTCGATA 3840
|||
QY 3781 GCGTCTCTGCTCTTCAGCATGTGAATGGGATCTCGGTCGATGTAAAAAATGTCGATA 3840

DB 3841 ATGTAAAAATGTCATGCGTCTTTTTCACACTTTCTGCACAAATGAATAGGGGAAATGT 3900
|||
QY 3841 ATGTAAAAATGTCATGCGTCTTTTTCACACTTTCTGCACAAATGAATAGGGGAAATGT 3900

DB 3901 ATTTAAATACATTTTGTATTTTTCACATTCACATGATTAACCCCATTTATTTTCGT 3960
|||
QY 3901 ATTTAAATACATTTTGTATTTTTCACATTCACATGATTAACCCCATTTATTTTCGT 3960

DB 3961 GAGCAACTTAAAAAGTAGAGATATTAGAGCGGAAAAACCAAAATTTCTCAAGATATTACC 4020
|||
QY 3961 GAGCAACTTAAAAAGTAGAGATATTAGAGCGGAAAAACCAAAATTTCTCAAGATATTACC 4020

DB 4021 TTTATTTGATTAATTATAGATGTTAATAAGCATATCTGAATGAAAGTCAGCAAAATATGT 4080
|||
QY 4021 TTTATTTGATTAATTATAGATGTTAATAAGCATATCTGAATGAAAGTCAGCAAAATATGT 4080

DB 4081 GCGAAACACCTGAAAAAATCAAAAATTTCTCGAAAAATTGAAAAAATGCAATTAATACA 4140
|||
QY 4081 GCGAAACACCTGAAAAAATCAAAAATTTCTCGAAAAATTGAAAAAATGCAATTAATACA 4140

DB 4141 TTTTTCATTTTCTACATCACATGATGTAGAAAAATTAAAAAGGAAATCAAAATTTCTA 4200
|||
QY 4141 TTTTTCATTTTCTACATCACATGATGTAGAAAAATTAAAAAGGAAATCAAAATTTCTA 4200

DB 4201 GAGCATATAATTGAATGAACATTCGGAATTTAAATGTGCGAAACGTCAAAAAAGAGA 4260
|||
QY 4201 GAGCATATAATTGAATGAACATTCGGAATTTAAATGTGCGAAACGTCAAAAAAGAGA 4260

DB 4261 AATTGGGTATCAAAATCGATCTTAAACCAACACATTTCAAGCATCCGCCACTCTTCAT 4320
|||
QY 4261 AATTGGGTATCAAAATCGATCTTAAACCAACACATTTCAAGCATCCGCCACTCTTCAT 4320

DB 4321 TCACCGGATGCTCTTCTCGGATACAGTTCAGTCTGTAATCGCTCATTCAGCAAAAGCTT 4380
|||
QY 4321 TCACCGGATGCTCTTCTCGGATACAGTTCAGTCTGTAATCGCTCATTCAGCAAAAGCTT 4380

DB 4381 CTGACCAACTCAATACATATTCATGAAGAGATATGAACCTTTGTGATGCACCAACCA 4440
|||
QY 4381 CTGACCAACTCAATACATATTCATGAAGAGATATGAACCTTTGTGATGCACCAACCA 4440

DB 4441 TAAGCCGTGTTTTCAGAGAGAAAAACCATGTACAGAACTTCTCGAGTCTCTGGAATGT 4500
|||
QY 4441 TAAGCCGTGTTTTCAGAGAGAAAAACCATGTACAGAACTTCTCGAGTCTCTGGAATGT 4500

DB 4501 GCCTCATATAATATGAACACTTTGAGCAGATGCCAACGGAATGTACCAAGCCG 4560
|||
QY 4501 GCCTCATATAATATGAACACTTTGAGCAGATGCCAACGGAATGTACCAAGCCG 4560

DB 4561 ACAAGGACAACTTTACCAATTTTTCAGATGCATGGGCTA 4600
|||
QY 4561 ACAAGGACAACTTTACCAATTTTTCAGATGCATGGGCTA 4600

RESULT 2
LOCUS CEC48D1 39908 bp DNA INV 23-NOV-1998
DEFINITION Caenorhabditis elegans cosmid C48D1, complete sequence.
ACCESSION Z81049
NID 91627677
VERSION 281049.1 GI:1627677
KEYWORDS HTG.
SOURCE
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 39908)
AUTHORS Burton, J.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-1996) Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwnematode.wustl.edu

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

Mpsrch_nn      n.a. - n.a. database search, using Smith-Waterman algorithm
Run on:        Sat Aug 7 18:20:15 1999;      MasPar time 1829.04 Seconds
Tabular output not generated.                1515.497 Million cell updates/sec

```

```

Title:                >US-08-287-669-18
Description:          (3601-4600) from US08287669.seq (5 of 10)
Perfect Score:       1000
N.A. Sequence:      3601 TCATGTGGTTGTTTCATATG.....TTGTCAGATGCATGGCGTA 4600
Comp:                AGTACACCAACAACAGTATAC.....AACAGTCTACGTACCCGAT

```

Scoring table: TABLE default

```
Nmatch      STD : Dbase 0; Query 0
```

Searched: 646147 seqs, 1385953633 bases x 2

```
Post-processing: Minimum Match 0%
Listing first 45 summaries
```

Database:

```

1:em_bal 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
7:em_in 8:em_om 9:em_or 10:em_ov 11:em_pat 12:em_ph
13:em_pl 14:em_ro 15:em_sts 16:em_vi
genbank111
Database:

```

17:gb_bal 18:gb_baz 19:gb_hgt1 20:gb_htg2 21:gb_in1
22:gb_in2 23:gb_cm 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_p11
28:gb_p12 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_v1

Statistics: Mean 11.159; Variance 6.687; scale 1.669

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	%	Length	DB	ID	Description	Pred. No.
	1	1000	100.0	7653	21	CELCED3A		Caenorhabditis elegans	0.00e+000
C	2	956	95.6	39908	21	CEC48D1		Caenorhabditis elegans	0.00e+000
C	3	938	93.8	314495	19	CEY67H2		Caenorhabditis elegans	0.00e+000
	4	154	15.4	39271	21	CEF52G2		Caenorhabditis elegans	4.67e-79
	5	154	15.4	110000	19	CEY45F10		Caenorhabditis elegans	4.67e-79
	6	154	15.4	256941	19	CEY62E10		Caenorhabditis elegans	4.67e-79
	7	142	14.2	39908	21	CEC48D1		Caenorhabditis elegans	5.60e-71
	8	142	14.2	314495	19	CEY67H2		Caenorhabditis elegans	5.60e-71
C	9	141	14.1	39271	21	CEF52G2		Caenorhabditis elegans	2.62e-70
C	10	141	14.1	110000	19	CEY45F10		Caenorhabditis elegans	2.62e-70
C	11	141	14.1	256941	19	CEY62E10		Caenorhabditis elegans	2.62e-70
C	12	127	12.7	7653	21	CELCED3A		Caenorhabditis elegans	5.87e-61
C	13	105	10.5	39752	22	CELC04F5		Caenorhabditis elegans	1.84e-46

C	14	94	9.4	39752	22	CELCO4F5	Caenorhabditis elegans	2.49e-39
C	15	88	8.8	37640	21	CEZK1073	Caenorhabditis elegans	1.75e-35
C	16	88	8.8	121290	19	CEH27C14	Caenorhabditis elegans	1.75e-35
C	17	83	8.3	309026	19	AC006560	Caenorhabditis elegans	2.65e-32
C	18	79	7.9	39553	21	CEUC8	Caenorhabditis elegans	8.87e-30
C	19	77	7.7	7218	25	I66494	Sequence 14 from paten	1.59e-28
C	20	76	7.6	37711	21	CEZK896	Caenorhabditis elegans	6.73e-28
C	21	76	7.6	309026	19	AC006760	Caenorhabditis elegans	6.73e-28
C	22	75	7.5	37059	21	CEK02E2	Caenorhabditis elegans	2.84e-27
C	23	75	7.5	110000	19	CEY102G3	Caenorhabditis elegans	2.84e-27
C	24	75	7.5	253812	19	CEY38H6	Caenorhabditis elegans	2.84e-27
C	25	73	7.3	38801	21	CEK09B11	Caenorhabditis elegans	4.98e-26
C	26	72	7.2	3747	19	AC006811	Caenorhabditis elegans	2.08e-25
C	27	72	7.2	34122	22	CELY66H1B	Caenorhabditis elegans	2.08e-25
C	28	72	7.2	317781	20	AC006906	Caenorhabditis elegans	2.08e-25
C	29	71	7.1	38681	22	CELR1IE3	Caenorhabditis elegans	8.63e-25
C	30	71	7.1	209365	19	AC006719	Caenorhabditis elegans	8.63e-25
C	31	70	7.0	31268	21	CEH08M01	Caenorhabditis elegans	3.58e-24
C	32	70	7.0	140702	20	AC006888	Caenorhabditis elegans	3.58e-24
C	33	69	6.9	30225	21	CEP29C12	Caenorhabditis elegans	1.48e-23
C	34	69	6.9	110000	19	CEY81G3	Caenorhabditis elegans	1.48e-23
C	35	69	6.9	267118	20	AC006889	Caenorhabditis elegans	1.48e-23
C	36	69	6.9	299202	19	AC006792	Caenorhabditis elegans	1.48e-23
C	37	68	6.8	29355	21	CEH13N06	Caenorhabditis elegans	6.07e-23
C	38	68	6.8	30225	21	CEP29C12	Caenorhabditis elegans	6.07e-23
C	39	67	6.7	37059	21	CEK02E2	Caenorhabditis elegans	2.49e-22
C	40	67	6.7	110000	19	CEY102G3	Caenorhabditis elegans	2.49e-22
C	41	67	6.7	253812	19	CEY38H6	Caenorhabditis elegans	2.49e-22
C	42	66	6.6	125590	21	CEY49E10	Caenorhabditis elegans	1.02e-21
C	43	63	6.3	45510	21	CEC32A3	Caenorhabditis elegans	6.74e-20
C	44	62	6.2	31430	21	CELC01C4	Caenorhabditis elegans	2.71e-19
C	45	61	6.1	22491	21	CEW07A8	Caenorhabditis elegans	1.08e-18

ALIGNMENTS

RESULT	1	CELCED3A	7653 bp	DNA	INV	23-FEB-1994
LOCUS						
DEFINITION		Caenorhabditis elegans cell death protein (ced-3) gene, complete cds.				
ACCESSION		L29052				
NID		g456416				
VERSION		L29052.1	GI:456416			
KEYWORDS		cell death protein; interleukin-1 beta converting enzyme.				
SOURCE		Caenorhabditis elegans (strain N2) DNA.				
ORGANISM		Caenorhabditis elegans				
		Eukaryotes; mitochondrial eukaryotes; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Caenorhabditis.				
REFERENCE		1 (bases 1 to 7653)				
AUTHORS		Yuan, J., Shaham, S., Ledoux, S., Ellis, H.M. and Horvitz, H.				
TITLE		The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1b-converting enzyme				
JOURNAL		Cell 75, 641-652 (1993)				
MEDLINE		94061982				
FEATURES						
Source		Location/Qualifiers				
		1..7653				
		/organism="Caenorhabditis elegans"				
		/strain="N2"				
		/db_xref="taxon:6239"				
	exon	2167..2366				
		/gene="ced-3"				
		/number=1				
	gene	join(2232..2366,2430..2575,2854..3107,4303..4634,5547..5760,5815..5942,6298..6537,7013..7075)				
		/gene="ced-3"				
		join(2232..2366,2430..2575,2854..3107,4303..4634,5547..5760,5815..5942,6298..6537,7013..7075)				
	CDS	/gene="ced-3"				
		/note="codes for a protein similar to mammalian interleukin-1b-converting enzyme"				
		/codon_start=1				
		/product="cell death protein"				

RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL NATURE 368:32-38(1994).

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA PAULEY A., GATTUNG S.,
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.,
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U53181; G1245690; -.
SQ SEQUENCE 134 AA; 14819 MW; 4F1CB795 CRC32;

Query Match 1.4%; Score 7; DB 5; Length 134;
Best Local Similarity 100.0%; Pred. No. 3.55e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 35 YSSPPVN 41
|||||
QY 166 YSSPPVN 172

RESULT 15
ID Q65580 PRELIMINARY; PRT; 162 AA.
AC Q65580;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE UL6 PROTEIN (FRAGMENT).
GN UL6.

OS BOVINE HERPESVIRUS TYPE 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SCHOENBOEKEN;
RX MEDLINE; 96135223.

RA SCHMITT J., KEIL G.M.;
RT "Identification and characterization of the bovine herpesvirus 1 UL7
RT gene and gene product which are not essential for virus replication
RT in cell culture.";
RL J. VIROL. 70:1091-1099(1996).
DR EMBL; X91751; G1006629; -.
FI NON_TER 1
SQ SEQUENCE 162 AA; 18656 MW; 4BA35B45 CRC32;

Query Match 1.4%; Score 7; DB 14; Length 162;
Best Local Similarity 100.0%; Pred. No. 3.55e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 100 RSRSRSR 106
|||||
Z, 150 RSRSRSR 156

Search completed: Tue Aug 10 11:41:55 1999
Job time : 109 secs.

AC 018749;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE PROTAMINE P1.
OS PLANIGALE SP. 1.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;
OC DASYROMORPHA; DASYURIDAE; PLANIGALE.
RN [1]
RP SEQUENCE FROM N.A.
RA KRAJEWSKI C., BLACKET M., BUCKLEY L., WESTERMAN M.;
RL MOL. PHYLOGENET. EVOL. 0:0-0(1997).
DR EMBL; AF001595; G2330935; -
SQ SEQUENCE 62 AA; 8290 MW; 12E28334 CRC32;

Query Match 1.4%; Score 7; DB 6; Length 62;
Best Local Similarity 100.0%; Pred. No. 3.55e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 9 RSRSR 15
QY 150 RSRSR 156

RESULT 11
ID 011288 PRELIMINARY; PRT; 128 AA.
AC 011288;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 13.2 KD PROTEIN (FRAGMENT).
GN X1-55.
OS MOLUSCUM CONTAGIOSUM VIRUS SUBTYPE 1 (MCV1).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC MOLUSCIPPOXVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RA MORATILLA M., AGROMAYOR M., NUNEZ A., FUNES J.M., VARAS A.J.,
RA LOPEZ-ESTEBARANZ J.L., ESTEBAN M., MARTIN-GALLARDO A.;
RL VIRUS GENES 0:0-0(0).
DR EMBL; U86881; G2105171; -
KW HYPOTHETICAL PROTEIN.
FT NON_TER 1
FT NON_TER 128
SQ SEQUENCE 128 AA; 13152 MW; 93E22F10 CRC32;

Query Match 1.4%; Score 7; DB 14; Length 128;
Best Local Similarity 100.0%; Pred. No. 3.55e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 4 RSRSR 10
QY 150 RSRSR 156

RESULT 12
ID 019225 PRELIMINARY; PRT; 130 AA.
AC 019225;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE F08H9.2 PROTEIN.
GN F08H9.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA WILD A.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR EMBL; 277657; E1345117; -
SQ SEQUENCE 130 AA; 14710 MW; EFBA5123 CRC32;

Query Match 1.4%; Score 7; DB 5; Length 130;
Best Local Similarity 100.0%; Pred. No. 3.55e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 32 RSRSR 38
QY 150 RSRSR 156

RESULT 13
ID 044177 PRELIMINARY; PRT; 134 AA.
AC 044177;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE ORF134.
OS SYNECHOCOCCUS PCC7002 PR-6.
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC 7002, ATCC 27264;
RA AKIYAMA H., KANAI S., HIRANO M., SUGIMOTO M., KIYOHARA M.;
RT "Cloning and characterization of Rubisco large subunit and small
RL subunit from Synechococcus sp. PCC7002.";
DR SUBMITTED (DEC-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; D13971; D1003583; -
SQ SEQUENCE 134 AA; 15269 MW; 6C6F48FF CRC32;

Query Match 1.4%; Score 7; DB 2; Length 134;
Best Local Similarity 100.0%; Pred. No. 3.55e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 102 RSRSLR 108
QY 6 RSRSLR 12

RESULT 14
ID 020104 PRELIMINARY; PRT; 134 AA.
AC 020104;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE THIS GENE MAY BEGIN IN THE NEXT COSMID.
GN F36D4.6.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL NATURE 368:32-38(1994).

RP [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA LANGSTON Z., WOHLDMANN P., GILLAM B.;
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 KA WATERSTON R.;
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; U97405; G2039375;
 SQ SEQUENCE 327 AA; 36425 MW; 81BF02D6 CRC32;

Query Match 1.6%; Score 8; DB 5; Length 327;
 Best Local Similarity 100.0%; Pred. No. 3.37e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 87 ARSRSR 94
 149 ARSRSR 156

RESULT 7
 ID 042840 PRELIMINARY; PRT; 456 AA.

AC 042840;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 46.7 KD PROTEIN (FRAGMENT).
 GN SPAC23A1.01C.
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
 OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
 OC SCHIZOSACCHAROMYCES.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972H-;
 RA MURPHY L., HARRIS D.;
 RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972H-;
 RA WOOD V., BARRELL B.G., RAJANDREAM M.A.;
 RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AL021813; E1250574; -
 KW HYPOTHETICAL PROTEIN.

FT NON_TER 456 456
 SQ SEQUENCE 456 AA; 46687 MW; 9216E7DF CRC32;

Query Match 1.6%; Score 8; DB 3; Length 456;
 Best Local Similarity 100.0%; Pred. No. 3.37e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 307 SSVSSTS 314
 134 SSVSSTS 141

RESULT 8
 ID 018207 PRELIMINARY; PRT; 619 AA.
 AC 018207;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)

DE Y48E1C.3 PROTEIN.
 GN Y48E1C.3.

OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.

RN [1]
 RP SEQUENCE FROM N.A.
 RA MCMURRAY A.;
 RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL NATURE 368:32-38(1994).
 DR EMBL; 293394; E1354641; -
 SQ SEQUENCE 619 AA; 72914 MW; 5A4A7D9C CRC32;

Query Match 1.6%; Score 8; DB 5; Length 619;
 Best Local Similarity 100.0%; Pred. No. 3.37e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 195 HLKVDL 202
 20 HLKVDL 27

RESULT 9
 ID 077435 PRELIMINARY; PRT; 1045 AA.

AC 077435;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE EG:34F3.4 PROTEIN.
 GN EG:34F3.4.
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.

RN [1]
 RP SEQUENCE FROM N.A.
 RA CATHERINE SALES, PHILIPPE VALENTI, ARETI DARLAMITSOV,
 RA NADINE HENDERSON, LORNA CAMPBELL, DAVID GLOVER;
 RT "Sequencing the distal X chromosome of Drosophila melanogaster.";
 RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]
 RP SEQUENCE FROM N.A.
 RA BENOS P.;
 RL SUBMITTED (OCT-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AL031583; E1321009; -
 SQ SEQUENCE 1045 AA; 116521 MW; 0EAD4D79 CRC32;

Query Match 1.6%; Score 8; DB 5; Length 1045;
 Best Local Similarity 100.0%; Pred. No. 3.37e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 129 THAKMDV 136
 447 THAKMDV 454

RESULT 10
 ID 018749 PRELIMINARY; PRT; 62 AA.

OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
 OC SCHIZOSACCHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA LYNE M., RAJANDREAM M.A., BARRELL B.G., WHITHEAD S., CHILLINGWORTH T.,
 RA CHURCHER C.M.;
 RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AL031158; E1313583; -
 FT NON_TER 174 174
 SQ SEQUENCE 174 AA; 20505 MW; 92ABDE3 CRC32;

Query Match 1.68; Score 8; DB 3; Length 174;
 Best Local Similarity 100.0%; Pred. No. 3.37e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

145 YSRARSRS 152
 146 YSRARSRS 153

RESULT 4
 ID 013854 PRELIMINARY; PRT; 263 AA.
 AC C13854;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE SPAC19G12.17C PROTEIN (FRAGMENT).
 GN SPAC19G12.17C.
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
 OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
 OC SCHIZOSACCHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA OLIVER K., HARRIS D.;
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA WOOD V., BARRELL B.G., RAJANDREAM M.A.;
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: Z97209; E1250596; -
 KW HYPOTHETICAL PROTEIN.
 FT NON_TER 1
 SQ SEQUENCE 263 AA; 26445 MW; 846167CF CRC32;

Query Match 1.68; Score 8; DB 3; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.37e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

8 SSVSSFTS 15
 134 SSVSSFTS 141

RESULT 5
 ID 007016 PRELIMINARY; PRT; 301 AA.
 AC 007016;
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 33.7 KD PROTEIN.
 GN YER.
 OS BACILLUS SUBTILIS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
 OC BACILLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA DENIZOT F.C.;
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE: 98044033.
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
 RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
 RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
 RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMINGS N.J., DANIEL R.A.,
 RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,
 RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
 RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
 RA GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
 RA GUISEPTI G., GUY B.J., HAGA K., HATECH J., HARWOOD C.R., HENAUT A.,
 RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
 RA JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
 RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
 RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
 RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
 RA NOONE D., O'REILLY M., OGAWA K., OGIMURA A., OUDEGA B., PARK S.H.,
 RA PARRO V., POHL T.M., PORTELELLA D., PORWOLLIK S., PRESCOTT A.M.,
 RA PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
 RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
 RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
 RA SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDI B.,
 RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
 RA TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
 RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
 RA VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
 RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 subtilis.";
 RL NATURE 390:249-256(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: Z94043; E133073; -
 DR EMBL: Z99121; E1186097; -
 DR PFAM: PF00005; ABC_tran; 1.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 301 AA; 33779 MW; C3824B26 CRC32;

Query Match 1.68; Score 8; DB 2; Length 301;
 Best Local Similarity 100.0%; Pred. No. 3.37e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

91 LKVDLE 98
 21 LKVDLE 28

RESULT 6
 ID 002155 PRELIMINARY; PRT; 327 AA.
 AC 002155;
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE CODED FOR BY C. ELEGANS CDNA YK77B5.5.
 GN T09B4.5.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEAE; RHABDITIA; RHABDITIDA;
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DUBBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,

970 6 1.2 2195 3 002822 VESICLE COAT PROTEIN S 2.08e+02
971 6 1.2 2244 14 089344 POLYMERASE. 2.08e+02
972 6 1.2 2272 5 017329 GAG, POL AND ENV PROTE 2.08e+02
973 6 1.2 2304 14 088893 RNAI POLYPROTEIN. 2.08e+02
974 6 1.2 2390 4 015020 BETA-SPECTRIN III. 2.08e+02
975 6 1.2 2478 13 093406 FIBRONECTIN. 2.08e+02
976 6 1.2 2489 3 006116 CHROMOSOME XVI COSMID 2.08e+02
977 6 1.2 2515 5 024551 TENASCIN-LIKE PROTEIN. 2.08e+02
978 6 1.2 2529 5 024605 ZINC FINGER PROTEIN. 2.08e+02
979 6 1.2 2606 5 021920 R1A8.7 PROTEIN. 2.08e+02
980 6 1.2 2731 5 018366 ODD OZ PRODUCT. 2.08e+02
981 6 1.2 2731 5 061307 ODD OZ PRODUCT. 2.08e+02
982 6 1.2 2761 5 018447 CODED FOR BY C. ELEGAN 2.08e+02
983 6 1.2 2893 2 025063 TOXIN-LIKE OUTER MEMBR 2.08e+02
984 6 1.2 3019 14 068801 POLYPROTEIN. 2.08e+02
985 6 1.2 3063 2 059497 FATTY-ACID SYNTHASE (E 2.08e+02
986 6 1.2 3066 11 062388 ATAXIA TELANGIECTASIA 2.08e+02
987 6 1.2 3175 14 089939 REPLICASE ORF1B POLYPR 2.08e+02
988 6 1.2 3413 2 054593 POLYKETIDE SYNTHASE. 2.08e+02
989 6 1.2 3722 2 094873 ALPHA-AMINOADIPYL-CYST 2.08e+02
990 6 1.2 3972 2 073139 HYPOTHETICAL 418.3 KD 2.08e+02
991 6 1.2 4470 14 039225 RNA-DIRECTED RNA POLYM 2.08e+02
992 6 1.2 4560 4 013787 ADOLIPROTEIN B100. 2.08e+02
993 6 1.2 4590 4 014517 CADHERIN-RELATED TUMOR 2.08e+02
994 6 1.2 4735 2 054666 POLYKETIDE SYNTHASE. 2.08e+02
995 6 1.2 4861 4 015751 P619. 2.08e+02
996 6 1.2 5069 2 052789 RIFAMYCIN POLYKETIDE S 2.08e+02
997 6 1.2 5071 5 091905 RYANODINE RECEPTOR. 2.08e+02
998 6 1.2 5107 5 094279 PARTIAL CDS. 2.08e+02
999 6 1.2 6420 2 095814 EK506 POLYKETIDE SYNTH 2.08e+02
1000 6 1.2 7829 5 018559 SIMILAR TO POLYKETIDE 2.08e+02

ALIGNMENTS

PRT; 495 AA.

RESULT 1
ID 002229 PRELIMINARY;
AC 002229;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE C48D1.2 PROTEIN.
GN C48D1.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA BURTON J.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
PP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
KA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
KA BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,
KA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
KA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
KA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
KA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
KA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
KA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
KA THERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
KA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
LR EMBL; 281049; E1344743; -.
DR PROSITE; PS01121; CASPASE_HIS; 1.
DR PROSITE; PS01122; CASPASE_CYS; 1.
SQ SEQUENCE 495 AA; 55549 MW; 7DC9ED9D CRC32;

Query Match 82.7%; Score 416; DB 5; Length 495;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;

Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 MRQDRSLLENNIMFSSHLKVDLEVLAKQVLNSDNGDMNSCGTVREKREIVKAV 60
QY 2 MRQDRSLLENNIMFSSHLKVDLEVLAKQVLNSDNGDMNSCGTVREKREIVKAV 61
Db 61 QRRGDVAFDAFYALRSTGHEGLAEVLEPLARSVDSDNAVEFECPMPSPASHRSRALSPAG 120
QY 62 QRRGDVAFDAFYALRSTGHEGLAEVLEPLARSVDSDNAVEFECPMPSPASHRSRALSPAG 121
Db 121 YTSPTRVHRDVSVSSTSYQDIYSRARSRSRSLHSSDRHNSPPVNAFSPQSSA 180
QY 122 YTSPTRVHRDVSVSSTSYQDIYSRARSRSRSLHSSDRHNSPPVNAFSPQSSA 181
Db 181 NSSFTGCCSLGYSSSRNRSFSGASGPTQYIFHEEDMNFVDAPTIISRYFDEKTYRNFSSP 240
QY 182 NSSFTGCCSLGYSSSRNRSFSGASGPTQYIFHEEDMNFVDAPTIISRYFDEKTYRNFSSP 241
Db 241 RGMCLINNEHEFQMPTRNGTKADKDNLTNFRMGYTVICKDNLTRGMLLTIRDFAKH 300
QY 242 RGMCLINNEHEFQMPTRNGTKADKDNLTNFRMGYTVICKDNLTRGMLLTIRDFAKH 301
Db 301 ESHGDSAILVILSHGEENVIIIGVDIPISHTHEIYDLLNAANAPRLANKPIVVOACRGE 360
QY 302 ESHGDSAILVILSHGEENVIIIGVDIPISHTHEIYDLLNAANAPRLANKPIVVOACRGE 361
Db 361 RRDNGFPVLDSVDGVPAPFLRGMWDRDGLFNLGCVRPQVQVWRKKPSQADILI 416
QY 362 RRDNGFPVLDSVDGVPAPFLRGMWDRDGLFNLGCVRPQVQVWRKKPSQADILI 417

RESULT 2
ID 079747 PRELIMINARY; PRT; 102 AA.
AC 079747;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CYTOCHROME B (FRAGMENT).
OS OMANOSAURA JAVAKARI.
OG MITOCHONDRION.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; LEPIDOSAURIA; SQUAMATA;
OC SCLEROGLOSSA; SCINCOMORPHA; LACERTOIDEA; LACERTIDAE; OMANOSAURA.
RN [1]
RP SEQUENCE FROM N.A.
RA HARRIS D.J.;
RT "Relationships of lacertid lizards (Reptilia: Lacertidae) estimated
RT from mitochondrial DNA sequences and morphology.";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF080351; G3560990; -.
KW MITOCHONDRION.
FT NON_TER 1 1
FT NON_TER 102 102
SQ SEQUENCE 102 AA; 11126 MW; 5B95367C CRC32;

Query Match 1.6%; Score 8; DB 8; Length 102;
Best Local Similarity 100.0%; Pred. No. 3.37e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 SPRGMCLI 10
QY 240 SPRGMCLI 247
RESULT 3
ID 074357 PRELIMINARY; PRT; 174 AA.
AC 074357;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POSSIBLE SPLICE FACTOR (FRAGMENT).
GN SPBC25D12.07C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;

824	6	1.2	845 14	036326	ENVELOPE PROTEIN.	2.08e+02	897	6	1.2	1028	2	068083	POTENTIAL OUTER MEMBRA	2.08e+02
825	6	1.2	845 14	036330	ENVELOPE PROTEIN.	2.08e+02	898	6	1.2	1053	5	024310	POLYPROTEIN.	2.08e+02
826	6	1.2	845 14	036329	ENVELOPE PROTEIN.	2.08e+02	899	6	1.2	1054	7	078109	MHC CLASS II TRANSACTI	2.08e+02
827	6	1.2	845 14	036315	ENVELOPE PROTEIN.	2.08e+02	900	6	1.2	1054	7	035482	HIGH MOLECULAR-WEIGHT	2.08e+02
828	6	1.2	845 14	036318	ENVELOPE PROTEIN.	2.08e+02	901	6	1.2	1086	4	013423	NICOTINAMIDE NUCLEOTID	2.08e+02
829	6	1.2	845 14	036317	ENVELOPE PROTEIN.	2.08e+02	902	6	1.2	1086	4	016796	NAD(P) TRANSHYDROGENAS	2.08e+02
830	6	1.2	845 14	036319	ENVELOPE PROTEIN.	2.08e+02	903	6	1.2	1094	14	016796	PUTATIVE VIRAL REPLIC	2.08e+02
831	6	1.2	845 14	036331	ENVELOPE PROTEIN.	2.08e+02	904	6	1.2	1095	4	060300	KIAA0553 PROTEIN (FRAG	2.08e+02
832	6	1.2	845 14	036328	ENVELOPE PROTEIN.	2.08e+02	905	6	1.2	1114	10	049345	PUTATIVE RNA HELICASE	2.08e+02
833	6	1.2	845 14	036327	ENVELOPE PROTEIN.	2.08e+02	906	6	1.2	1116	14	098632	RNA BINDING PROTEIN.	2.08e+02
834	6	1.2	845 14	042056	ENVELOPE PROTEIN.	2.08e+02	907	6	1.2	1144	5	062179	F21H7.9 PROTEIN.	2.08e+02
835	6	1.2	846 5	017897	C10C5.1 PROTEIN.	2.08e+02	908	6	1.2	1148	14	055519	127K MINOR OUTER CAPSI	2.08e+02
836	6	1.2	849 14	066731	ENVELOPE PROTEIN (FRAG	2.08e+02	909	6	1.2	1152	5	045463	F36D3.5 PROTEIN.	2.08e+02
837	6	1.2	852 11	088943	POTASSIUM CHANNEL.	2.08e+02	910	6	1.2	1155	3	014239	HYPOTHETICAL 128.3 KD	2.08e+02
838	6	1.2	853 14	089469	ENVELOPE POLYPROTEIN.	2.08e+02	911	6	1.2	1157	5	027089	PYRUVATE:FERREDOXIN OX	2.08e+02
839	6	1.2	854 4	075580	NEURONAL DELAYED-RECTI	2.08e+02	912	6	1.2	1157	5	027088	PYRUVATE:FERREDOXIN OX	2.08e+02
840	6	1.2	855 3	042722	STEL2ALPHA.	2.08e+02	913	6	1.2	1163	4	092618	MYELOBLAST KIAA0222.	2.08e+02
841	6	1.2	858 14	089473	ENVELOPE POLYPROTEIN.	2.08e+02	914	6	1.2	1170	5	016587	F21E9.1 PROTEIN.	2.08e+02
842	6	1.2	858 14	047968	DNA GYRASE A SUBUNIT.	2.08e+02	915	6	1.2	1173	3	013313	MYOSIN-LIKE PROTEIN SP	2.08e+02
843	6	1.2	859 14	036321	ENVELOPE PROTEIN.	2.08e+02	916	6	1.2	1174	6	095168	TIGHT JUNCTION PROTEIN	2.08e+02
844	6	1.2	859 14	036325	ENVELOPE PROTEIN.	2.08e+02	917	6	1.2	1182	11	035595	PTCH2.	2.08e+02
845	6	1.2	859 14	036323	ENVELOPE PROTEIN.	2.08e+02	918	6	1.2	1186	5	021227	K04G2.8A PROTEIN.	2.08e+02
846	6	1.2	859 14	036324	ENVELOPE PROTEIN.	2.08e+02	919	6	1.2	1188	5	062302	K04G2.8B PROTEIN.	2.08e+02
847	6	1.2	859 14	036350	ENVELOPE PROTEIN.	2.08e+02	920	6	1.2	1192	5	061853	F33E11.3 PROTEIN.	2.08e+02
848	6	1.2	859 14	036351	ENVELOPE PROTEIN.	2.08e+02	921	6	1.2	1195	1	058524	HYPOTHETICAL PROTEIN M	2.08e+02
849	6	1.2	859 14	036316	ENVELOPE PROTEIN.	2.08e+02	922	6	1.2	1198	2	053730	POLYKETIDE SYNTHASE (F	2.08e+02
850	6	1.2	859 14	092814	ENVELOPE POLYPROTEIN.	2.08e+02	923	6	1.2	1199	5	091349	SIMILARITY TO SHORT RE	2.08e+02
851	6	1.2	859 14	036356	ENVELOPE PROTEIN.	2.08e+02	924	6	1.2	1205	14	056300	IE-A PROTEIN.	2.08e+02
852	6	1.2	859 14	036320	ENVELOPE PROTEIN.	2.08e+02	925	6	1.2	1213	3	059801	PUTATIVE HELICASE.	2.08e+02
853	6	1.2	859 14	036322	ENVELOPE PROTEIN.	2.08e+02	926	6	1.2	1214	5	025338	DELTA-LATROINSECTOTOXI	2.08e+02
854	6	1.2	859 14	036334	ENVELOPE PROTEIN.	2.08e+02	927	6	1.2	1232	1	058318	HYPOTHETICAL PROTEIN M	2.08e+02
855	6	1.2	859 14	036335	ENVELOPE PROTEIN.	2.08e+02	928	6	1.2	1238	5	018780	SIMILARITY TO MOUSE SM	2.08e+02
856	6	1.2	859 14	036336	ENVELOPE PROTEIN.	2.08e+02	929	6	1.2	1259	8	035058	ATPA INTRON2 ORF.	2.08e+02
857	6	1.2	859 14	036337	ENVELOPE PROTEIN.	2.08e+02	930	6	1.2	1262	5	020771	F54D5.5.	2.08e+02
858	6	1.2	859 14	036338	ENVELOPE PROTEIN.	2.08e+02	931	6	1.2	1275	4	015057	KIAA0349 (FRAGMENT).	2.08e+02
859	6	1.2	859 14	036353	ENVELOPE PROTEIN.	2.08e+02	932	6	1.2	1277	11	035821	PAR INTERACTING PROTEI	2.08e+02
860	6	1.2	859 14	036355	ENVELOPE PROTEIN.	2.08e+02	933	6	1.2	1282	11	060520	PAIRED AMPHIPATHIC HEL	2.08e+02
861	6	1.2	859 14	036354	ENVELOPE PROTEIN.	2.08e+02	934	6	1.2	1287	3	006047	ANTIVIRAL PROTEIN (SMI	2.08e+02
862	6	1.2	859 14	042060	ENVELOPE PROTEIN.	2.08e+02	935	6	1.2	1291	10	038800	COL-O PUTATIVE RNA HEL	2.08e+02
863	6	1.2	859 14	099268	ENV POLYPROTEIN PRECUR	2.08e+02	936	6	1.2	1300	5	077261	EG:115C2.10 PROTEIN.	2.08e+02
864	6	1.2	859 14	036339	ENVELOPE PROTEIN.	2.08e+02	937	6	1.2	1302	14	065741	RNA-DIRECTED RNA POLYM	2.08e+02
865	6	1.2	859 5	021401	COSMID K09E3.	2.08e+02	938	6	1.2	1315	3	074431	MEMBRANE ATPASE.	2.08e+02
866	6	1.2	860 14	036333	ENVELOPE PROTEIN.	2.08e+02	939	6	1.2	1344	11	035851	P160 MYB-BINDING PROTE	2.08e+02
867	6	1.2	867 5	044235	HRSH2.	2.08e+02	940	6	1.2	1364	4	075092	MEGF2 (FRAGMENT).	2.08e+02
868	6	1.2	868 10	038710	ABIETADIENE CYCLASE.	2.08e+02	941	6	1.2	1371	2	073337	SENSORY TRANSDUCTION H	2.08e+02
869	6	1.2	869 2	085041	CARBOXYISOME SHELL POLY	2.08e+02	942	6	1.2	1378	11	061138	PATERNALLY EXPRESSED P	2.08e+02
870	6	1.2	871 5	017088	ZC308.1 PROTEIN.	2.08e+02	943	6	1.2	1381	14	066628	CAPSID PROTEIN.	2.08e+02
871	6	1.2	880 3	015895	ATP-BINDING PROTEIN.	2.08e+02	944	6	1.2	1463	14	069298	MAJOR IMMEDIATE EARLY	2.08e+02
872	6	1.2	884 5	006163	CHROMOSOME XII COSMID	2.08e+02	945	6	1.2	1503	13	073677	MYELOBLAST KIAA0234.	2.08e+02
873	6	1.2	884 7	029675	MHC CLASS II TRANSACTI	2.08e+02	946	6	1.2	1503	13	073677	CYSTIC FIBROSIS TRANSM	2.08e+02
874	6	1.2	897 11	070495	PLENTY-OF-PROLINES-101	2.08e+02	947	6	1.2	1509	11	061194	PHOSPHATIDYLINOSITOL 3	2.08e+02
875	6	1.2	899 11	063527	RETINOBLASTOMA PROTEIN	2.08e+02	948	6	1.2	1510	5	022699	SIMILARITY TO MYOSIN H	2.08e+02
876	6	1.2	902 8	035063	COXI INTRON1 ORF.	2.08e+02	949	6	1.2	1536	3	006625	PROLIFERATION POTENTI	2.08e+02
877	6	1.2	902 2	007686	ORF A PROTEIN.	2.08e+02	950	6	1.2	1551	4	014160	ZINC FINGER PROTEIN.	2.08e+02
878	6	1.2	912 14	090278	ENVELOPE PROTEIN.	2.08e+02	951	6	1.2	1560	11	097868	KIAA0147 PROTEIN (FRAG	2.08e+02
879	6	1.2	925 5	020099	SIMILAR TO DILUTE MYOS	2.08e+02	952	6	1.2	1571	11	054978	PROLIFERATION POTENTI	2.08e+02
880	6	1.2	943 4	014664	PSD-95/SAP90-ASSOCIATE	2.08e+02	953	6	1.2	1602	5	017679	COSG5.6 (FRAGMENT).	2.08e+02
881	6	1.2	946 2	070888	(BCTA).	2.08e+02	954	6	1.2	1635	5	061168	F48A11.1 PROTEIN.	2.08e+02
882	6	1.2	957 5	019204	SIMILAR TO THROMBOSPON	2.08e+02	955	6	1.2	1658	11	017383	PHOSPHATIDYLINOSITOL 3	2.08e+02
883	6	1.2	959 10	023291	KINESIN HOMOLOG.	2.08e+02	956	6	1.2	1661	5	027095	RNA POLYMERASE II.	2.08e+02
884	6	1.2	959 10	022987	UNKNOWN PROTEIN.	2.08e+02	957	6	1.2	1676	10	023332	CENTROMERE PROTEIN HOM	2.08e+02
885	6	1.2	969 5	027502	K03A11.1 PROTEIN.	2.08e+02	958	6	1.2	1686	4	000443	PHOSPHOINOSITIDE 3-KIN	2.08e+02
886	6	1.2	975 5	091357	SIMILARITY TO ARABIDOP	2.08e+02	959	6	1.2	1698	5	024440	(CDNA1) PROTEIN 4.1 HO	2.08e+02
887	6	1.2	979 2	053609	PUTATIVE MEMBRANE PROT	2.08e+02	960	6	1.2	1742	5	024463	PROJECTIN (FRAGMENT).	2.08e+02
888	6	1.2	994 5	077088	DORSAL B.	2.08e+02	961	6	1.2	1770	2	084419	PUTATIVE OUTER MEMBRAN	2.08e+02
889	6	1.2	997 3	074325	HYPOTHETICAL 111.3 KD	2.08e+02	962	6	1.2	1834	14	066233	POLYPROTEIN.	2.08e+02
890	6	1.2	1001 11	088664	SERINE/THREONINE PROTE	2.08e+02	963	6	1.2	1839	5	017383	HUM-2.	2.08e+02
891	6	1.2	1003 5	001915	SIMILARITY TO N-CHIMAE	2.08e+02	964	6	1.2	1865	10	081909	T7123.15 PROTEIN.	2.08e+02
892	6	1.2	1005 1	058718	HYPOTHETICAL PROTEIN M	2.08e+02	965	6	1.2	1932	5	001483	COSMID C06A5.	2.08e+02
893	6	1.2	1017 11	089048	ELK CHANNEL 1.	2.08e+02	966	6	1.2	1937	2	030482	PKS MODULE 4.	2.08e+02
894	6	1.2	1026 3	006315	CHROMOSOME XII COSMID	2.08e+02	967	6	1.2	2042	5	025766	3D7VAR1 (FRAGMENT).	2.08e+02
895	6	1.2	1028 11	097528	NB-3.	2.08e+02	968	6	1.2	2115	4	014980	NUMA PROTEIN.	2.08e+02
896	6	1.2	1028 4	015042	KIAA0332 (FRAGMENT).	2.08e+02	969	6	1.2	2140	3	002316	LE.MFB1 PROTEIN.	2.08e+02

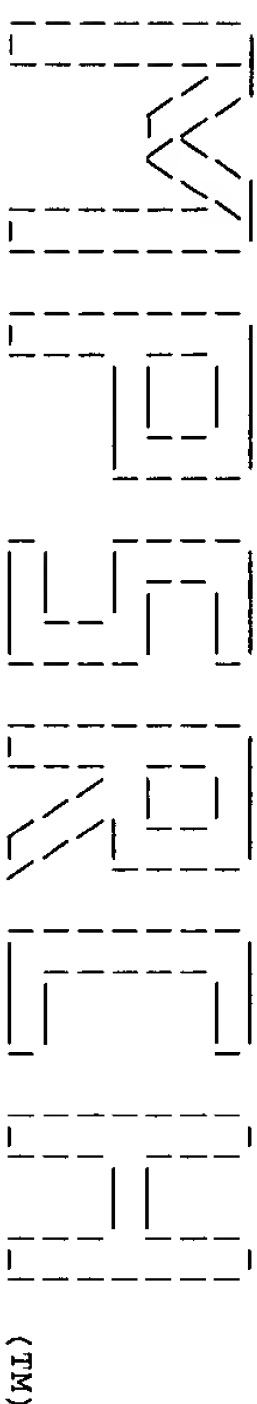
678	6	1.2	566 14	Q98095	HEMAGGLUTININ.	2.08e+02	751	6	1.2	680 11	070583	MIDLINE 1 (MIDLINE 1 P	2.08e+02
679	6	1.2	566 14	Q82774	HAEMAGGLUTININ PRECURS	2.08e+02	752	6	1.2	684 1	058969	HYPOTHETICAL HELICASE	2.08e+02
680	6	1.2	566 14	Q82503	HEMAGGLUTININ.	2.08e+02	753	6	1.2	689 13	073620	NUCLEAR PROTEIN SDK2.	2.08e+02
681	6	1.2	566 14	P88836	HAEMAGGLUTININ HA (FRA	2.08e+02	754	6	1.2	695 5	P91908	M04C7.2 PROTEIN.	2.08e+02
682	6	1.2	566 4	Q16716	PYRUVATE KINASE (EC 2.	2.08e+02	755	6	1.2	697 14	Q87026	CAPSID.	2.08e+02
683	6	1.2	566 14	Q82770	HEMAGGLUTININ PRECURSO	2.08e+02	756	6	1.2	699 5	Q19640	COSMID F20D12.	2.08e+02
684	6	1.2	566 5	Q21740	ROSD1.8 PROTEIN.	2.08e+02	757	6	1.2	700 13	Q91453	STONUSTOXIN BETA-SUBUN	2.08e+02
685	6	1.2	566 14	Q67164	HEMAGGLUTININ.	2.08e+02	758	6	1.2	701 2	051403	NA+/H+ ANTIPORTER (NAP	2.08e+02
686	6	1.2	566 14	Q82832	HEMAGGLUTININ PRECURSO	2.08e+02	759	6	1.2	701 14	Q65568	HYPOTHETICAL 72.6 KD P	2.08e+02
687	6	1.2	566 14	Q84110	HEMAGGLUTININ.	2.08e+02	760	6	1.2	703 6	P79122	PININ.	2.08e+02
688	6	1.2	566 14	Q82500	HEMAGGLUTININ PRECURSO	2.08e+02	761	6	1.2	706 3	Q12296	CHROMOSOME XV READING	2.08e+02
689	6	1.2	566 14	Q67043	HEMAGGLUTININ PRECURSO	2.08e+02	762	6	1.2	706 1	052000	ORE H1130.	2.08e+02
690	6	1.2	566 14	Q67042	SIMILAR TO CYTOPLASMIC	2.08e+02	763	6	1.2	706 10	081324	F6N15.3 PROTEIN.	2.08e+02
691	6	1.2	568 5	Q21728	TRBL.	2.08e+02	764	6	1.2	708 13	Q98993	VERRUCOTOXIN ALPHA PRE	2.08e+02
692	6	1.2	572 2	P71187	HYPOTHETICAL 64.5 KD P	2.08e+02	765	6	1.2	708 5	Q22961	SIMILARITY TO TPR DOMA	2.08e+02
693	6	1.2	572 3	P87231	VELVET A.	2.08e+02	766	6	1.2	711 4	060899	NUCLEAR PROTEIN SDK3 (2.08e+02
694	6	1.2	573 3	Q74625	PYROPHOSPHATE--FRUCTOS	2.08e+02	767	6	1.2	712 4	075853	WUGSC:H.DJ0808A01.2 PR	2.08e+02
695	6	1.2	573 2	Q83553	AXI 1-LIKE PROTEIN.	2.08e+02	768	6	1.2	718 5	077136	TRANSLATION INITIATION	2.08e+02
696	6	1.2	576 14	Q64884	HEMAGGLUTININ (FRAGMEN	2.08e+02	769	6	1.2	718 5	073619	PUTATIVE TRANSMEMBRANE	2.08e+02
697	6	1.2	576 14	Q86677	TYROSINE HYDROXYLASE T	2.08e+02	770	6	1.2	720 14	036332	NUCLEAR PROTEIN.	2.08e+02
698	6	1.2	579 5	Q24000	TRANSCRIPTONAL REPRES	2.08e+02	771	6	1.2	721 13	Q91902	ENVELOPE PROTEIN.	2.08e+02
699	6	1.2	579 5	Q76734	TCTEX-3.	2.08e+02	772	6	1.2	725 11	035691	X-DELTA-1.	2.08e+02
700	6	1.2	580 11	Q54808	DOLICHOLO-P-GLUCOSE SYN	2.08e+02	773	6	1.2	725 4	043671	PCF11P HOMOLOG.	2.08e+02
701	6	1.2	581 1	Q29674	CELL DIVISION PROTEIN	2.08e+02	774	6	1.2	727 2	P96583	PROBABLE DNA TOPOISOME	2.08e+02
702	6	1.2	581 2	Q69074	CHROMOSOME XII READING	2.08e+02	775	6	1.2	729 2	Q55545	HYPOTHETICAL 84.2 KD P	2.08e+02
703	6	1.2	583 3	Q12372	PGRS-FAMILY PROTEIN.	2.08e+02	776	6	1.2	731 2	032913	MALATE SYNTHASE G.	2.08e+02
704	6	1.2	584 2	Q53575	PYRUVATE KINASE (EC 2.	2.08e+02	777	6	1.2	734 2	Q32043	GTP PYROPHOSPHOKINASE	2.08e+02
705	6	1.2	587 4	Q16715	FIDIPIDINE.	2.08e+02	778	6	1.2	737 5	Q45746	T05A12.4 PROTEIN.	2.08e+02
706	6	1.2	588 5	Q77293	ATAXIA-TELANGIECTASIA	2.08e+02	779	6	1.2	737 3	Q45746	PLASMODI RP4 TRAE.	2.08e+02
707	6	1.2	588 4	Q14134	CELL DIVISION PROTEIN	2.08e+02	780	6	1.2	738 2	Q60215	P1067 PROTEIN.	2.08e+02
708	6	1.2	590 2	Q69075	PENICILLIN BINDING PRO	2.08e+02	781	6	1.2	739 3	Q13658	VPI LIKE PROTEIN (FRAG	2.08e+02
709	6	1.2	595 2	Q67249	SIMILAR TO KINESIN-LIK	2.08e+02	782	6	1.2	739 10	Q65420	FAS BINDING PROTEIN (D	2.08e+02
710	6	1.2	595 5	Q19745	CDC25-REGULATORY KINAS	2.08e+02	783	6	1.2	740 4	Q15141	FAS-BINDING PROTEIN DA	2.08e+02
711	6	1.2	598 13	P70032	COSMID SSSD1.	2.08e+02	784	6	1.2	740 4	Q14747	PININ.	2.08e+02
712	6	1.2	599 5	Q22040	PYRUVATE KINASE (EC 2.	2.08e+02	785	6	1.2	743 5	Q22859	W04D2.6 PROTEIN.	2.08e+02
713	6	1.2	599 4	Q75758	SIMILAR TO GANCICLOVIR	2.08e+02	786	6	1.2	744 5	Q46076	1-EVIDENCE-PREDICTED B	2.08e+02
714	6	1.2	600 14	Q67670	T-BOX TRANSCRIPTION FA	2.08e+02	787	6	1.2	748 2	P96527	ORE10.	2.08e+02
715	6	1.2	602 11	Q70306	POLO LIKE KINASE.	2.08e+02	788	6	1.2	748 14	Q98200	MC032L.	2.08e+02
716	6	1.2	603 11	Q62673	EXCINUCLEASE ABC, SUBU	2.08e+02	789	6	1.2	748 12	Q62534	SERINE/THREONINE PROTE	2.08e+02
717	6	1.2	603 2	Q51413	PUTATIVE KP78 PROTEIN	2.08e+02	790	6	1.2	752 5	Q62530	CARS-CYP (EC 5.2.1.8).	2.08e+02
718	6	1.2	604 5	Q17049	GH3 LIKE PROTEIN.	2.08e+02	791	6	1.2	754 4	Q13427	IKB KINASE BETA.	2.08e+02
719	6	1.2	612 10	Q81829	GAG-LIKE PROTEIN.	2.08e+02	792	6	1.2	757 11	Q88351	PROBABLE NA(+)/H(+) AN	2.08e+02
720	6	1.2	613 10	Q64427	TOM70 GENE.	2.08e+02	793	6	1.2	759 3	Q14123	BETA-GLUCOSIDASE.	2.08e+02
721	6	1.2	614 3	Q13499	TYPE I RESTRICTION MOD	2.08e+02	794	6	1.2	764 5	Q20977	F58E6.1 PROTEIN.	2.08e+02
722	6	1.2	616 1	Q27025	PUTATIVE ARGINYLT-TRNA	2.08e+02	795	6	1.2	765 2	Q54183	CHITINASE.	2.08e+02
723	6	1.2	618 3	Q74781	BAND 4.1-LIKE PROTEIN	2.08e+02	796	6	1.2	765 2	Q54183	HYPOTHETICAL 87.5 KD P	2.08e+02
724	6	1.2	619 13	Q57457	CAPSID-ASSOCIATED PROT	2.08e+02	797	6	1.2	769 3	P79149	PININ.	2.08e+02
725	6	1.2	624 14	Q41446	JM1 PROTEIN, COMPLETE	2.08e+02	798	6	1.2	773 6	P87050	PROBABLE SERINE/THREON	2.08e+02
726	6	1.2	627 4	Q60826	L3169.3.	2.08e+02	799	6	1.2	775 3	Q86228	VP4 PROTEIN.	2.08e+02
727	6	1.2	627 5	Q15832	TELOMERASE.	2.08e+02	800	6	1.2	776 14	Q86229	VP4 PROTEIN.	2.08e+02
728	6	1.2	631 9	Q37967	GLUCOSAMINE-FRUCTOSE-6	2.08e+02	801	6	1.2	776 14	Q86229	VP4 PROTEIN.	2.08e+02
729	6	1.2	635 2	Q83833	PRIMASE.	2.08e+02	802	6	1.2	778 3	Q83894	TEX PROTEIN (TEX).	2.08e+02
730	6	1.2	636 2	Q52200	(-)-4S-LIMONENE SYNTHA	2.08e+02	803	6	1.2	779 3	Q06224	CHROMOSOME XII COSMID	2.08e+02
731	6	1.2	637 10	Q22340	DNAX.	2.08e+02	804	6	1.2	790 3	Q07732	CHROMOSOME IV READING	2.08e+02
732	6	1.2	637 2	Q32464	Y48E1B.13 PROTEIN.	2.08e+02	805	6	1.2	792 5	Q61840	F56D6.1 PROTEIN (FRAGM	2.08e+02
733	6	1.2	642 5	Q18203	TRANSCRIPTION INITIATI	2.08e+02	806	6	1.2	801 4	Q13436	HISTONE TRANSCRIPTIONA	2.08e+02
734	6	1.2	642 3	Q74319	C11G6.1 PROTEIN.	2.08e+02	807	6	1.2	807 3	Q74309	UL47H.	2.08e+02
735	6	1.2	643 5	Q17907	COSMID F14D12.	2.08e+02	808	6	1.2	808 14	Q69320	HYPOTHETICAL 91.3 KD P	2.08e+02
736	6	1.2	645 5	Q19448	SIMILARITY TO EXOZ GEN	2.08e+02	809	6	1.2	811 10	Q49385	T22H6.6 PROTEIN.	2.08e+02
737	6	1.2	645 5	Q21688	F53H1.1 PROTEIN.	2.08e+02	810	6	1.2	813 5	Q22691	C2-HC TYPE ZINC FINGER	2.08e+02
738	6	1.2	646 5	Q45076	BIP/GRP78 PRECURSOR.	2.08e+02	811	6	1.2	815 11	P70589	NK-TUMOR RECOGNITION M	2.08e+02
739	6	1.2	653 5	Q26733	HYPOTHETICAL 70.8 KD P	2.08e+02	812	6	1.2	816 5	Q43273	CODED FOR BY C. ELEGAN	2.08e+02
740	6	1.2	657 2	Q07192	GLABRA2 HOMOLOG.	2.08e+02	813	6	1.2	817 1	Q22291	816AA LONG HYPOTHETICA	2.08e+02
741	6	1.2	661 10	Q23611	BETA-GALACTOSIDASE (EC	2.08e+02	814	6	1.2	817 3	Q13665	P1074 PROTEIN.	2.08e+02
742	6	1.2	662 6	Q62800	ACETYL-COA SYNTHETASE	2.08e+02	815	6	1.2	820 4	Q60585	SER/ARG-RELATED NUCLEA	2.08e+02
743	6	1.2	662 2	Q25686	INAD PROTEIN.	2.08e+02	816	6	1.2	830 14	Q83863	95.1KD PUTATIVE NONSTR	2.08e+02
744	6	1.2	665 5	Q23823	HYPOTHETICAL 75.3 KD P	2.08e+02	817	6	1.2	836 11	P97359	TAF195.	2.08e+02
745	6	1.2	667 2	Q06075	PUTATIVE TRANSCRIPTION	2.08e+02	818	6	1.2	837 5	Q76751	PUTATIVE ZINC METALLOP	2.08e+02
746	6	1.2	667 4	Q15344	ATP-DEPENDENT HELICASE	2.08e+02	819	6	1.2	840 11	Q88556	CONDUCTIN.	2.08e+02
747	6	1.2	668 2	Q83538	PHOSPHOENOLPYRUVATE CA	2.08e+02	820	6	1.2	841 2	Q85709	RELA PROTEIN.	2.08e+02
748	6	1.2	669 10	Q04914	T14P8.3 PROTEIN.	2.08e+02	821	6	1.2	842 10	Q49527	POLLEN-SPECIFIC PROTEI	2.08e+02
749	6	1.2	674 10	Q81292	WACLAW.	2.08e+02	822	6	1.2	844 4	Q43796	VOLTAGE GATED POTASSIU	2.08e+02
750	6	1.2	679 5	Q61347			823	6	1.2				

532	6	1.2	401	5	045907	NHR-65 PROTEIN.	2.08e+02	605	6	1.2	483	14	041517	55 KDA PROTEIN.	2.08e+02
533	6	1.2	404	13	090420	47 KDA HEAT SHOCK PROT	2.08e+02	606	6	1.2	487	2	050901	BETA-1,4-GLYCANASE.	2.08e+02
534	6	1.2	407	13	P70053	PAX-6.	2.08e+02	607	6	1.2	489	10	065518	HYPOTHETICAL 54.1 KD P	2.08e+02
535	6	1.2	408	2	007796	POSSIBLE TRANSPOSASE.	2.08e+02	608	6	1.2	490	2	P95098	HYPOTHETICAL 54.0 KD P	2.08e+02
536	6	1.2	409	3	012128	YOR3320W FROM CHROMOSO	2.08e+02	609	6	1.2	491	10	049660	PREDICTED PROTEIN.	2.08e+02
537	6	1.2	410	13	090323	SERINE PROTEASE INHIBI	2.08e+02	610	6	1.2	492	13	093532	OTOKERATIN.	2.08e+02
538	6	1.2	412	11	063556	SERINE PROTEASE INHIBI	2.08e+02	611	6	1.2	494	5	P91653	SERENDIPITY (FRAGMENT)	2.08e+02
539	6	1.2	413	5	062467	Y45F10B.2 PROTEIN.	2.08e+02	612	6	1.2	496	4	Q15414	YRMI PROTEIN.	2.08e+02
540	6	1.2	416	2	066645	BRAIN AND MUSCLE ARNT-	2.08e+02	613	6	1.2	496	2	025600	CONSERVED HYPOTHETICAL	2.08e+02
541	6	1.2	417	2	025751	HYPOTHETICAL 47.8 KD P	2.08e+02	614	6	1.2	497	5	P91708	SERENDIPITY (FRAGMENT)	2.08e+02
542	6	1.2	417	5	023229	COLICIN TOLERANCE-LIKE	2.08e+02	615	6	1.2	498	5	P91709	SERENDIPITY (FRAGMENT)	2.08e+02
543	6	1.2	417	11	035624	PRODUCT OF THE 2.2KB T	2.08e+02	616	6	1.2	499	4	Q13675	ALPHA 1C ADRENERGIC RE	2.08e+02
544	6	1.2	417	11	025624	IPL1 AND AURORA RELATE	2.08e+02	617	6	1.2	500	11	088962	STEROL 12-ALPHA HYDROX	2.08e+02
545	6	1.2	418	4	099413	CONSERVED HYPOTHETICAL	2.08e+02	618	6	1.2	500	11	002766	STEROL 12-ALPHA HYDROX	2.08e+02
546	6	1.2	420	14	067008	PAX6 (FRAGMENT).	2.08e+02	619	6	1.2	501	11	035074	PROSTACYCLIN SYNTHASE	2.08e+02
547	6	1.2	420	14	067008	HYPOTHETICAL 46.1 KD P	2.08e+02	620	6	1.2	501	11	075004	CD24P.	2.08e+02
548	6	1.2	421	13	006832	POLYPROTEIN PRECURSOR	2.08e+02	621	6	1.2	501	11	062969	PROSTACYCLIN SYNTHASE.	2.08e+02
549	6	1.2	421	13	P70001	XENOPUS PAX-6 LONG (FR	2.08e+02	622	6	1.2	505	10	064667	F22013.18.	2.08e+02
550	6	1.2	421	10	049409	PROTEIN KINASE - LIKE	2.08e+02	623	6	1.2	508	10	082263	F17A22.35 PROTEIN.	2.08e+02
551	6	1.2	422	11	035669	FLICE PARTIAL CDS (FLI	2.08e+02	624	6	1.2	509	8	047547	CYTCHROME C OXIDASE S	2.08e+02
552	6	1.2	422	5	025122	HRO-CDC25.	2.08e+02	625	6	1.2	509	10	Q40870	LEGUMIN-LIKE STORAGE P	2.08e+02
553	6	1.2	423	2	P95746	HYPOTHETICAL 44.1 KD P	2.08e+02	626	6	1.2	510	2	Q59913	RNA POLYMERASE SIGMA F	2.08e+02
554	6	1.2	423	10	082626	ERG PROTEIN.	2.08e+02	627	6	1.2	511	14	011450	ENVELOPE GLYCOPROTEIN.	2.08e+02
555	6	1.2	424	1	026914	CONSERVED PROTEIN.	2.08e+02	628	6	1.2	511	5	016512	T05H4.7 PROTEIN.	2.08e+02
556	6	1.2	425	2	025142	FUCOSYLTRANSFERASE.	2.08e+02	629	6	1.2	513	2	067052	F32D1.9 PROTEIN.	2.08e+02
557	6	1.2	425	11	070290	G-PROTEIN-COUPLED INWA	2.08e+02	630	6	1.2	513	5	016293	HISTIDINE KINASE SENSO	2.08e+02
558	6	1.2	427	4	000148	NUCLEAR RNA HELICASE.	2.08e+02	631	6	1.2	515	14	093015	ENVELOPE GLYCOPROTEIN	2.08e+02
559	6	1.2	427	5	Q22491	COSMID T14F9.	2.08e+02	632	6	1.2	515	3	065519	REQUIRED FOR VACUOLE S	2.08e+02
560	6	1.2	428	2	033959	TYL ORFX.	2.08e+02	633	6	1.2	517	10	065519	HYPOTHETICAL 58.5 KD P	2.08e+02
561	6	1.2	429	4	Q13729	ALPHA 1C ADRENERGIC RE	2.08e+02	634	6	1.2	520	5	Q93877	MEMBRANE TRANSPORTER D	2.08e+02
562	6	1.2	432	14	009639	HEMAGGLUTININ PRECURSO	2.08e+02	635	6	1.2	521	10	022848	CYP3CIP.	2.08e+02
563	6	1.2	433	2	051871	POLY(3-HYDROXYBUTYRATE	2.08e+02	636	6	1.2	521	10	048926	F55A8.2 PROTEIN (FRAGM	2.08e+02
564	6	1.2	436	13	042348	PAX6 PROTEIN.	2.08e+02	637	6	1.2	521	5	076360	HSMP-1.	2.08e+02
565	6	1.2	437	14	009638	HEMAGGLUTININ PRECURSO	2.08e+02	638	6	1.2	523	4	Q99932	RNA POLYMERASE SIGMA F	2.08e+02
566	6	1.2	437	10	064810	PUTATIVE SERINE CARBOX	2.08e+02	639	6	1.2	528	2	Q59814	GP64/67-EFP-ACMNPV ORF	2.08e+02
567	6	1.2	440	2	083691	HYPOTHETICAL 47.7 KD P	2.08e+02	640	6	1.2	530	14	092483	C.	2.08e+02
568	6	1.2	441	14	Q07499	NUCLEOCAPSID PROTEIN.	2.08e+02	641	6	1.2	530	5	018064	HYPOHETICAL 60.6 KD P	2.08e+02
569	6	1.2	442	10	023335	RNA HELICASE HOMOLOG.	2.08e+02	642	6	1.2	531	14	009642	F21A3.5 PROTEIN.	2.08e+02
570	6	1.2	443	2	087843	HYPOTHETICAL 48.1 KD P	2.08e+02	643	6	1.2	531	2	055524	HYPOHETICAL 60.6 KD P	2.08e+02
571	6	1.2	445	10	080503	F16B22.13 PROTEIN.	2.08e+02	644	6	1.2	532	5	017828	PUTATIVE CD98 PROTEIN.	2.08e+02
572	6	1.2	451	14	092492	ME53-ACMNPV ORF139.	2.08e+02	645	6	1.2	533	11	055221	HOMEBOX-CONTAINING PR	2.08e+02
573	6	1.2	451	5	Q20648	CODED FOR BY C. ELEGAN	2.08e+02	646	6	1.2	534	5	043989	DNA POLYMERASE SIGMA F	2.08e+02
574	6	1.2	451	5	Q19321	COSMID F10G7.	2.08e+02	647	6	1.2	534	14	066643	HYPOTHETICAL 62.0 KD P	2.08e+02
575	6	1.2	451	8	Q13882	RIBULOSE BIPHOSPHATE	2.08e+02	648	6	1.2	535	3	042653	HYPOHETICAL 62.0 KD P	2.08e+02
576	6	1.2	451	4	Q13882	TYROSINE KINASE.	2.08e+02	649	6	1.2	535	10	049600	IMPORTIN ALPHA-LIKE PR	2.08e+02
577	6	1.2	453	5	P92013	R10D12.15 PROTEIN.	2.08e+02	650	6	1.2	538	11	088900	CD25 B-TYPE TYROSINE	2.08e+02
578	6	1.2	453	13	Q91886	PAIRED-TYPE HOMODOMAI	2.08e+02	651	6	1.2	539	10	081053	T18E12.12 PROTEIN.	2.08e+02
579	6	1.2	454	4	Q14060	COPROPORPHYRINGEN OXI	2.08e+02	652	6	1.2	543	5	Q21064	CYTOSOLMIC INTERMEDIA	2.08e+02
580	6	1.2	455	4	060451	ALPHA 1A ADRENERGIC RE	2.08e+02	653	6	1.2	543	5	Q19286	SIMILAR TO INTERMEDIAT	2.08e+02
581	6	1.2	458	10	022803	PUTATIVE SERINE CARBOX	2.08e+02	654	6	1.2	546	3	Q08904	CHROMOSOME XV READING	2.08e+02
582	6	1.2	459	2	Q49101	OUTER MEMBRANE PROTEIN	2.08e+02	655	6	1.2	549	5	Q19287	SIMILAR TO INTERMEDIAT	2.08e+02
583	6	1.2	462	2	P95644	RNA POLYMERASE SIGMA F	2.08e+02	656	6	1.2	552	10	Q96336	AMP-BINDING PROTEIN.	2.08e+02
584	6	1.2	462	2	075958	STEROL 12-ALPHA HYDROX	2.08e+02	657	6	1.2	552	10	Q96336	AMP-BINDING PROTEIN.	2.08e+02
585	6	1.2	464	4	014676	APOTOTIC CASPASE MCHS	2.08e+02	658	6	1.2	553	10	P93837	PYRUVATE DEHYDROGENASE	2.08e+02
586	6	1.2	465	2	006664	RNA POLYMERASE BETA' S	2.08e+02	659	6	1.2	553	1	Q26576	3-PHOSPHOINOSITIDE DEP	2.08e+02
587	6	1.2	466	11	054913	ALPHA 1A-ADRENERGIC RE	2.08e+02	660	6	1.2	553	10	081808	T-BOX 14 (MMTBX14).	2.08e+02
588	6	1.2	467	3	036014	PROBABLE VACUOLAR AMIN	2.08e+02	661	6	1.2	556	4	015530	HEMAGGLUTININ (FRAGMEN	2.08e+02
589	6	1.2	467	3	074430	TRIGLYCERIDE LIPASE-CH	2.08e+02	662	6	1.2	559	11	055173	HEMAGGLUTININ PRECURSO	2.08e+02
590	6	1.2	470	2	P95121	HYPOTHETICAL 46.7 KD P	2.08e+02	663	6	1.2	559	11	054840	PUTATIVE IMMUNITY.	2.08e+02
591	6	1.2	471	5	P90988	COSMID B0432.	2.08e+02	664	6	1.2	561	14	082764	HYPOTHETICAL 62.5 KD P	2.08e+02
592	6	1.2	471	1	028786	HYPOTHETICAL 54.2 KD P	2.08e+02	665	6	1.2	562	14	082768	HYPOHETICAL 63.6 KD P	2.08e+02
593	6	1.2	471	5	Q20178	F38H4.3 PROTEIN.	2.08e+02	666	6	1.2	562	2	086284	STEROL 12-ALPHA HYDROX	2.08e+02
594	6	1.2	472	2	Q44072	PARACRYSTALLINE SURFAC	2.08e+02	667	6	1.2	563	10	065520	PROSTACYCLIN SYNTHASE.	2.08e+02
595	6	1.2	476	2	025366	FUCOSYLTRANSFERASE.	2.08e+02	668	6	1.2	563	10	049520	F22013.18.	2.08e+02
596	6	1.2	477	5	026218	CYTCHROME C OXIDASE S	2.08e+02	669	6	1.2	565	14	Q82766	CYTCHROME C OXIDASE S	2.08e+02
597	6	1.2	477	8	063695	CYTCHROME C OXIDASE S	2.08e+02	670	6	1.2	565	10	048774	LEGUMIN-LIKE STORAGE P	2.08e+02
598	6	1.2	478	5	Q25684	MAJOR MERZOITE SURFAC	2.08e+02	671	6	1.2	566	14	067039	ENVELOPE GLYCOPROTEIN.	2.08e+02
599	6	1.2	478	11	062905	VITRONECTIN.	2.08e+02	672	6	1.2	566	14	067040	T05H4.7 PROTEIN.	2.08e+02
600	6	1.2	480	2	Q30511	ALPHA 1,3-FUCOSYLTRANSF	2.08e+02	673	6	1.2	566	14	082834	F32D1.9 PROTEIN.	2.08e+02
601	6	1.2	480	11	089110	CASPASE-8.	2.08e+02	674	6	1.2	566	14	Q67007	ENVELOPE GLYCOPROTEIN	2.08e+02
602	6	1.2	480	5	061839	F56D6.2 PROTEIN.	2.08e+02	675	6	1.2	566	14	Q98092	REQUIRED FOR VACUOLE S	2.08e+02
603	6	1.2	480	5	001159	D2089.1 PROTEIN.	2.08e+02	676	6	1.2	566	14	Q98094	MEMBRANE TRANSPORTER D	2.08e+02
604	6	1.2	480	14	Q69466	GLYCOPROTEIN C PRECURS	2.08e+02	677	6	1.2	566	14	Q98093	CYP3CIP.	2.08e+02

386	6	1.2	344 14	007767	HEMAGGLUTININ PRECURSO	2.08e+02	459	6	1.2	361 7	046881	MHC CLASS I DLA-64.	2.08e+02
387	6	1.2	344 14	007753	HEMAGGLUTININ PRECURSO	2.08e+02	460	6	1.2	364 14	068800	NS5 (FRAGMENT)	2.08e+02
388	6	1.2	344 14	007774	HEMAGGLUTININ PRECURSO	2.08e+02	461	6	1.2	364 14	068816	NS5 (FRAGMENT)	2.08e+02
389	6	1.2	344 14	067060	HAEMAGGLUTININ (FRAGME	2.08e+02	462	6	1.2	364 14	068795	NS5 (FRAGMENT)	2.08e+02
390	6	1.2	344 14	067080	HAEMAGGLUTININ (FRAGME	2.08e+02	463	6	1.2	364 14	068775	NS5 (FRAGMENT)	2.08e+02
391	6	1.2	344 14	067066	HAEMAGGLUTININ (FRAGME	2.08e+02	464	6	1.2	364 14	068815	NS5 (FRAGMENT)	2.08e+02
392	6	1.2	344 14	067063	HAEMAGGLUTININ (FRAGME	2.08e+02	465	6	1.2	364 14	081553	NS5 (FRAGMENT)	2.08e+02
393	6	1.2	344 14	067082	HAEMAGGLUTININ (FRAGME	2.08e+02	466	6	1.2	364 14	068789	NS5 (FRAGMENT)	2.08e+02
394	6	1.2	344 14	067077	HAEMAGGLUTININ (FRAGME	2.08e+02	467	6	1.2	364 14	068809	NS5 (FRAGMENT)	2.08e+02
395	6	1.2	344 14	067075	HAEMAGGLUTININ (FRAGME	2.08e+02	468	6	1.2	364 14	068807	NS5 (FRAGMENT)	2.08e+02
396	6	1.2	344 14	010623	HEMAGGLUTININ (FRAGME	2.08e+02	469	6	1.2	364 6	077572	X-LINKED VISUAL PIGMEN	2.08e+02
397	6	1.2	344 14	067109	HAEMAGGLUTININ PRECURS	2.08e+02	470	6	1.2	364 6	077576	X-LINKED VISUAL PIGMEN	2.08e+02
398	6	1.2	344 14	067069	HAEMAGGLUTININ (FRAGME	2.08e+02	471	6	1.2	364 6	077575	X-LINKED VISUAL PIGMEN	2.08e+02
399	6	1.2	344 14	067083	HAEMAGGLUTININ (FRAGME	2.08e+02	472	6	1.2	364 2	044471	PUTATIVE TARRATE DEHY	2.08e+02
400	6	1.2	344 14	067084	HAEMAGGLUTININ (FRAGME	2.08e+02	473	6	1.2	364 6	077569	X-LINKED VISUAL PIGMEN	2.08e+02
401	6	1.2	344 14	067129	HAEMAGGLUTININ PRECURS	2.08e+02	474	6	1.2	364 6	062799	LONG-WAVELENGTH SENSIT	2.08e+02
402	6	1.2	344 14	067100	HAEMAGGLUTININ PRECURS	2.08e+02	475	6	1.2	364 6	062860	COLOR VISION OPSIN.	2.08e+02
403	6	1.2	344 14	067096	HAEMAGGLUTININ PRECURS	2.08e+02	476	6	1.2	364 6	077571	X-LINKED VISUAL PIGMEN	2.08e+02
404	6	1.2	344 14	067114	HAEMAGGLUTININ PRECURS	2.08e+02	477	6	1.2	364 6	077570	X-LINKED VISUAL PIGMEN	2.08e+02
405	6	1.2	344 14	067124	HAEMAGGLUTININ PRECURS	2.08e+02	478	6	1.2	364 6	077573	X-LINKED VISUAL PIGMEN	2.08e+02
406	6	1.2	344 14	067121	HAEMAGGLUTININ PRECURS	2.08e+02	479	6	1.2	364 6	077574	X-LINKED VISUAL PIGMEN	2.08e+02
407	6	1.2	344 14	067076	HAEMAGGLUTININ PRECURS	2.08e+02	480	6	1.2	366 5	016891	F13A2.6 PROTEIN.	2.08e+02
408	6	1.2	344 14	067130	HAEMAGGLUTININ (FRAGME	2.08e+02	481	6	1.2	367 14	041165	A683L PROTEIN.	2.08e+02
409	6	1.2	344 14	067068	HAEMAGGLUTININ (FRAGME	2.08e+02	482	6	1.2	368 11	097353	GDP-L-FUCOSE: BETA-D-G	2.08e+02
410	6	1.2	344 14	067067	HAEMAGGLUTININ (FRAGME	2.08e+02	483	6	1.2	369 2	069143	PUTATIVE PYRUVATE-FERR	2.08e+02
411	6	1.2	344 14	067061	HAEMAGGLUTININ (FRAGME	2.08e+02	484	6	1.2	370 13	P70002	XENOPUS PAX-6 SHORT (F	2.08e+02
412	6	1.2	344 14	067101	HAEMAGGLUTININ PRECURS	2.08e+02	485	6	1.2	370 10	022940	UNKNOWN PROTEIN.	2.08e+02
413	6	1.2	344 14	067127	HAEMAGGLUTININ PRECURS	2.08e+02	486	6	1.2	370 14	009637	HEMAGGLUTININ PRECURSO	2.08e+02
414	6	1.2	344 14	067065	HAEMAGGLUTININ (FRAGME	2.08e+02	487	6	1.2	370 14	009636	HEMAGGLUTININ PRECURSO	2.08e+02
415	6	1.2	344 14	007776	HEMAGGLUTININ PRECURSO	2.08e+02	488	6	1.2	371 10	081734	HYPOTHETICAL 40.8 KD P	2.08e+02
416	6	1.2	344 14	007780	HEMAGGLUTININ PRECURSO	2.08e+02	489	6	1.2	371 2	085595	ALANINE DEHYDROGENASE	2.08e+02
417	6	1.2	344 14	007775	HEMAGGLUTININ PRECURSO	2.08e+02	490	6	1.2	377 4	075381	PEROXISOMAL MEMBRANE A	2.08e+02
418	6	1.2	344 14	007770	HEMAGGLUTININ PRECURSO	2.08e+02	491	6	1.2	377 4	043534	NATURAL KILLER CELL IN	2.08e+02
419	6	1.2	345 7	095468	HEMAGGLUTININ PRECURSO	2.08e+02	492	6	1.2	377 4	099706	NATURAL KILLER CELL RE	2.08e+02
420	6	1.2	347 13	059756	MHC CLASS I ANTIGEN (F	2.08e+02	493	6	1.2	377 4	099560	NK RECEPTOR.	2.08e+02
421	6	1.2	347 13	057605	HYPOTHETICAL 39.0 KD P	2.08e+02	494	6	1.2	377 4	014621	KILLER CELL RECEPTOR.	2.08e+02
422	6	1.2	349 5	021602	UV-SENSITIVE CONE OPSI	2.08e+02	495	6	1.2	377 4	075601	EVOLUTIONARILY RELATED	2.08e+02
423	6	1.2	349 5	021602	M88.4 PROTEIN.	2.08e+02	496	6	1.2	379 1	029504	HYPOTHETICAL 42.7 KD P	2.08e+02
424	6	1.2	349 4	060399	CGM6_HUMAN.	2.08e+02	497	6	1.2	379 8	048364	CYTCHROME B.	2.08e+02
425	6	1.2	350 2	007564	HYPOTHETICAL 39.3 KD P	2.08e+02	498	6	1.2	379 8	047992	CYTCHROME B.	2.08e+02
426	6	1.2	351 14	082485	NORTH AMERICAN SWINE I	2.08e+02	499	6	1.2	379 8	047959	CYTCHROME B.	2.08e+02
427	6	1.2	351 14	082486	NORTH AMERICAN SWINE I	2.08e+02	500	6	1.2	379 1	054623	SIMILAR TO SULOLOBUS S	2.08e+02
428	6	1.2	351 14	012677	HEMAGGLUTININ (FRAGMEN	2.08e+02	501	6	1.2	380 8	047976	CYTCHROME B.	2.08e+02
429	6	1.2	352 5	019220	F08G5.1 PROTEIN.	2.08e+02	502	6	1.2	380 8	047961	CYTCHROME B.	2.08e+02
430	6	1.2	352 10	022443	SEED COAT PEROXIDASE P	2.08e+02	503	6	1.2	382 9	021998	GALACTOKINASE.	2.08e+02
431	6	1.2	352 2	030869	HAETI RESTRICTION ENZY	2.08e+02	504	6	1.2	382 8	034340	CYTCHROME B LIGHT STR	2.08e+02
432	6	1.2	353 5	021051	COSMID F59G1.	2.08e+02	505	6	1.2	382 8	035558	CYTCHROME B LIGHT STR	2.08e+02
433	6	1.2	353 14	067223	HA PRECURSOR (FRAGMENT	2.08e+02	506	6	1.2	382 14	035515	CYTCHROME B LIGHT STR	2.08e+02
434	6	1.2	354 13	087499	MEL-1C(A) MELATONIN RE	2.08e+02	507	6	1.2	382 14	035518	CYTCHROME B LIGHT STR	2.08e+02
435	6	1.2	354 1	028281	IRON-SULFUR BINDING RE	2.08e+02	508	6	1.2	383 14	067012	VIRUS NUCLEOPROTEIN.	2.08e+02
436	6	1.2	354 14	067335	INFLUENZA A/SL/2/87, A	2.08e+02	509	6	1.2	383 5	045938	HEMAGGLUTININ (FRAGMEN	2.08e+02
437	6	1.2	354 13	087496	MEL-1C(B) MELATONIN RE	2.08e+02	510	6	1.2	384 14	066931	Y44F5A.1 PROTEIN.	2.08e+02
438	6	1.2	355 1	029857	CONSERVED HYPOTHETICAL	2.08e+02	511	6	1.2	384 14	069357	HOMOLOGUE OF HSV-1 GI.	2.08e+02
439	6	1.2	356 2	055216	CARMINOMYCIN 4-O-METHY	2.08e+02	512	6	1.2	384 14	066931	HYPOTHETICAL 42.3 KD P	2.08e+02
440	6	1.2	356 1	028044	HYPOTHETICAL 40.9 KD P	2.08e+02	513	6	1.2	385 14	067005	GI GENE PRECURSOR.	2.08e+02
441	6	1.2	357 6	018872	KERATINOCYTE GROWTH FA	2.08e+02	514	6	1.2	385 14	009641	POLYPROTEIN PRECURSOR	2.08e+02
442	6	1.2	357 14	067091	HEMAGGLUTININ PRECURSO	2.08e+02	515	6	1.2	387 2	095416	HEMAGGLUTININ PRECURSO	2.08e+02
443	6	1.2	357 2	056948	PESTICIN.	2.08e+02	516	6	1.2	388 2	088027	NIRJ.	2.08e+02
444	6	1.2	357 14	067090	HEMAGGLUTININ PRECURSO	2.08e+02	517	6	1.2	388 14	065360	HYPOTHETICAL 41.6 KD P	2.08e+02
445	6	1.2	357 2	057159	PESTICIN.	2.08e+02	518	6	1.2	388 14	065360	HYPOTHETICAL 42.1 KD P	2.08e+02
446	6	1.2	358 2	005918	HYPOTHETICAL 37.8 KD P	2.08e+02	519	6	1.2	390 5	017274	T27A1.4 PROTEIN.	2.08e+02
447	6	1.2	358 2	069073	TRANSPOSASE.	2.08e+02	520	6	1.2	391 5	044148	C49A9.7 PROTEIN.	2.08e+02
448	6	1.2	358 14	067088	HEMAGGLUTININ PRECURSO	2.08e+02	521	6	1.2	391 14	089925	PROBABLY SITE-SPECIFIC	2.08e+02
449	6	1.2	358 2	056916	PUTATIVE TWO-COMPONENT	2.08e+02	522	6	1.2	392 14	039170	HAEMAGGLUTININ (FRAGME	2.08e+02
450	6	1.2	358 2	069960	GREEN OPSIN (GREEN-SEN	2.08e+02	523	6	1.2	392 14	039173	HAEMAGGLUTININ (FRAGME	2.08e+02
451	6	1.2	359 11	035476	ENVELOPE PROTEIN (FRAG	2.08e+02	524	6	1.2	393 2	045305	HAEMAGGLUTININ (FRAGME	2.08e+02
452	6	1.2	360 14	093066	NK RECEPTOR.	2.08e+02	525	6	1.2	393 5	053405	C47A10.3 PROTEIN.	2.08e+02
453	6	1.2	360 4	099559	HEMAGGLUTININ PRECURSO	2.08e+02	526	6	1.2	395 11	023440	HYPOHETICAL 40.7 KD P	2.08e+02
454	6	1.2	360 4	009640	CHROMOSOME XVI READING	2.08e+02	527	6	1.2	395 3	002340	SERINE/THREONINE KINAS	2.08e+02
455	6	1.2	360 14	008930	MHC CLASS I DLA-88.	2.08e+02	528	6	1.2	395 3	002341	ACETOHYDROXY-ACID ISOM	2.08e+02
456	6	1.2	361 7	046882	FISSION YEAST (FRAGMEN	2.08e+02	529	6	1.2	396 2	055794	HYPOTHETICAL 44.6 KD P	2.08e+02
457	6	1.2	361 3	P78808		2.08e+02	530	6	1.2	397 2	P74590	HYPOTHETICAL 43.5 KD P	2.08e+02
458	6	1.2	361 3	P78808		2.08e+02	531	6	1.2	400 2	088009	HYPOTHETICAL 40.7 KD P	2.08e+02

240	6	1.2	232	2	084263	DNA POL III EPSILON CH	2.08e+02	313	6	1.2	311	2	034859	YKOV PROTEIN.	2.08e+02
241	6	1.2	233	9	048422	P PROTEIN.	2.08e+02	314	6	1.2	312	2	033744	HYPOTHETICAL 33.8 KD P	2.08e+02
242	6	1.2	233	2	032558	ORF 1 PROTEIN.	2.08e+02	315	6	1.2	313	10	080500	F16B22.9 PROTEIN.	2.08e+02
243	6	1.2	233	2	P71005	HYPOTHETICAL 26.3 KD P	2.08e+02	316	6	1.2	313	11	088310	INTELECTIN.	2.08e+02
244	6	1.2	236	2	P76942	PUTATIVE PROPHAGE SF6-	2.08e+02	317	6	1.2	318	4	04875	RNA BINDING PROTEIN.	2.08e+02
245	6	1.2	237	10	023436	SIMILARITY TO HYPOTHET	2.08e+02	318	6	1.2	319	10	040780	HOMEOBOX-LEUCINE ZIPPE	2.08e+02
246	6	1.2	243	11	088205	TESTIS SPECIFIC PROTEI	2.08e+02	319	6	1.2	319	2	057316	KILB PROTEIN.	2.08e+02
247	6	1.2	243	2	083006	PHOTOSYNTHETIC REACTIO	2.08e+02	320	6	1.2	319	2	053165	PROTEIN INVOLVED WITH	2.08e+02
248	6	1.2	244	2	005934	PUTATIVE REGULATORY PR	2.08e+02	321	6	1.2	320	1	058990	320A LONG HYPOTHETICA	2.08e+02
249	6	1.2	245	4	000547	CL-9.	2.08e+02	322	6	1.2	322	6	046560	PYRIDOXAL KINASE (EC 2	2.08e+02
250	6	1.2	246	2	054045	PUTATIVE CHAPERONE.	2.08e+02	323	6	1.2	323	5	018035	SIMILAR TO YEAST MAK16	2.08e+02
251	6	1.2	246	3	074648	ISOAMYL ACETATE HYDROL	2.08e+02	324	6	1.2	324	1	028153	TRANSCRIPTIONAL REGULA	2.08e+02
252	6	1.2	247	1	029253	HYPOTHETICAL 28.2 KD P	2.08e+02	325	6	1.2	324	10	041097	VEGETATIVE STORAGE PRO	2.08e+02
253	6	1.2	248	2	005437	HYPOTHETICAL 27.1 KD P	2.08e+02	326	6	1.2	325	5	019060	HA1 CHAIN.	2.08e+02
254	6	1.2	249	2	P74414	HYPOTHETICAL 27.3 KD P	2.08e+02	327	6	1.2	325	14	083964	HEMAGGLUTININ (FRAGME	2.08e+02
255	6	1.2	249	9	080074	ANTI REPRESSOR.	2.08e+02	328	6	1.2	325	14	067115	HEMAGGLUTININ (FRAGME	2.08e+02
256	6	1.2	251	2	034449	YOOD PROTEIN.	2.08e+02	329	6	1.2	326	14	067116	HEMAGGLUTININ (FRAGME	2.08e+02
257	6	1.2	251	9	064113	PUTATIVE ANTIREPRESSOR	2.08e+02	330	6	1.2	326	5	045738	TO3E6.5 PROTEIN.	2.08e+02
258	6	1.2	251	11	054895	GUANYLATE KINASE MEMBR	2.08e+02	331	6	1.2	327	14	009645	HAEMAGGLUTININ (FRAGME	2.08e+02
259	6	1.2	252	1	026647	CONSERVED PROTEIN.	2.08e+02	332	6	1.2	327	14	009645	HEMAGGLUTININ (FRAGME	2.08e+02
260	6	1.2	252	5	016726	TO7D3.2 PROTEIN.	2.08e+02	333	6	1.2	327	14	009649	HEMAGGLUTININ (FRAGME	2.08e+02
261	6	1.2	253	2	055668	HYPOTHETICAL 27.3 KD P	2.08e+02	334	6	1.2	327	14	009648	HEMAGGLUTININ (FRAGME	2.08e+02
262	6	1.2	253	5	020665	SIMILAR TO PHOSPHOLIPA	2.08e+02	335	6	1.2	327	14	009646	HEMAGGLUTININ (FRAGME	2.08e+02
263	6	1.2	256	5	022208	SIMILARITY TO GLYCOPRO	2.08e+02	336	6	1.2	327	14	009647	HEMAGGLUTININ (FRAGME	2.08e+02
264	6	1.2	257	11	089094	CASPASE-14.	2.08e+02	337	6	1.2	327	14	009650	HEMAGGLUTININ (FRAGME	2.08e+02
265	6	1.2	258	4	014400	GLUTAMATE DEHYDROGENAS	2.08e+02	338	6	1.2	327	14	009644	HEMAGGLUTININ (FRAGME	2.08e+02
266	6	1.2	259	5	062221	F33H2.8 PROTEIN.	2.08e+02	339	6	1.2	327	14	009651	HEMAGGLUTININ (FRAGME	2.08e+02
267	6	1.2	261	2	P72737	HYPOTHETICAL 29.5 KD P	2.08e+02	340	6	1.2	327	14	009613	HEMAGGLUTININ (FRAGME	2.08e+02
268	6	1.2	262	11	088468	TLS-ASSOCIATED PROTEIN	2.08e+02	341	6	1.2	327	14	009613	HEMAGGLUTININ (FRAGME	2.08e+02
269	6	1.2	262	4	075494	TLS-ASSOCIATED PROTEIN	2.08e+02	342	6	1.2	327	5	093345	HEMAGGLUTININ (FRAGME	2.08e+02
270	6	1.2	262	5	093783	F59F4.4 PROTEIN.	2.08e+02	343	6	1.2	327	14	009652	C36B1.11 PROTEIN.	2.08e+02
271	6	1.2	263	13	090470	HOMXD1.	2.08e+02	344	6	1.2	328	14	039837	HEMAGGLUTININ (FRAGME	2.08e+02
272	6	1.2	267	5	018233	SIMILARITY WITH WILMS'	2.08e+02	345	6	1.2	328	14	039837	HEMAGGLUTININ (FRAGME	2.08e+02
273	6	1.2	267	3	001138	CHITIN SYNTHASE (FRAGM	2.08e+02	346	6	1.2	332	1	027009	HEMAGGLUTININ (FRAGME	2.08e+02
274	6	1.2	268	2	083851	ABC TRANSPORTER, ATP-B	2.08e+02	347	6	1.2	333	2	054063	TUNGSTEN FORMYLMETHANO	2.08e+02
275	6	1.2	269	3	029751	MOLYBDOPTERIN OXIDORED	2.08e+02	348	6	1.2	333	2	032631	SPAN.	2.08e+02
276	6	1.2	275	1	074305	PUTATIVE SPLICING PROT	2.08e+02	349	6	1.2	334	10	P95113	ALPHA-(1,3)-FUCCOSYLTRA	2.08e+02
277	6	1.2	278	2	053864	HYPOTHETICAL 30.1 KD P	2.08e+02	350	6	1.2	334	10	P94088	HYPOTHETICAL 34.0 KD P	2.08e+02
278	6	1.2	280	5	068106	COBALF TRANSPORT ATP-B	2.08e+02	351	6	1.2	337	14	011317	PUTATIVE ASPARTATE-ARG	2.08e+02
279	6	1.2	281	2	023120	W02B12.2 PROTEIN.	2.08e+02	352	6	1.2	338	5	044340	SIMILAR TO VAROLA E2L	2.08e+02
280	6	1.2	281	5	083580	HYPOTHETICAL 31.0 KD P	2.08e+02	353	6	1.2	338	14	066664	L-LACTATE DEHYDROGENAS	2.08e+02
281	6	1.2	282	5	016874	C13A2.4 PROTEIN.	2.08e+02	354	6	1.2	340	2	048759	CAPSID PROTEIN.	2.08e+02
282	6	1.2	283	13	093417	CASPASE-3.	2.08e+02	355	6	1.2	342	4	014622	CLPC ATPASE (MEC).	2.08e+02
283	6	1.2	283	10	042784	NEUTROPHIL PROTEIN (FR	2.08e+02	356	6	1.2	342	2	051379	KILR CELL RECEPTOR.	2.08e+02
284	6	1.2	284	6	099331	LIPASE ACTIVATOR PROTE	2.08e+02	357	6	1.2	343	4	043803	HYPOTHETICAL 38.8 KD P	2.08e+02
285	6	1.2	284	11	062222	PAIRED BOX PROTEIN 6 (2.08e+02	358	6	1.2	343	14	082833	NK RECEPTOR.	2.08e+02
286	6	1.2	290	2	P70999	HYPOTHETICAL 32.4 KD P	2.08e+02	359	6	1.2	344	14	067137	HEMAGGLUTININ (FRAGME	2.08e+02
287	6	1.2	293	2	045076	HYDROXYQUINOL 1,2-DIOX	2.08e+02	360	6	1.2	344	14	084074	INFLUENZA A/SINGAPORE/	2.08e+02
288	6	1.2	293	10	038696	(CLONE KFXET-I) XET PR	2.08e+02	361	6	1.2	344	14	084075	INFLUENZA A/TAIWAN/1/8	2.08e+02
289	6	1.2	294	10	081290	T14P8.21 PROTEIN.	2.08e+02	362	6	1.2	344	14	067054	HEMAGGLUTININ (FRAGME	2.08e+02
290	6	1.2	295	14	086360	G PROTEIN (FRAGMENT).	2.08e+02	363	6	1.2	344	14	067057	HEMAGGLUTININ (FRAGME	2.08e+02
291	6	1.2	295	6	077621	ALPHA-1A ADRENORECEPTO	2.08e+02	364	6	1.2	344	14	067072	HEMAGGLUTININ (FRAGME	2.08e+02
292	6	1.2	295	2	066147	HYPOTHETICAL 31.0 KD P	2.08e+02	365	6	1.2	344	14	067058	HEMAGGLUTININ (FRAGME	2.08e+02
293	6	1.2	297	14	091947	ATTACHMENT PROTEIN.	2.08e+02	366	6	1.2	344	14	067064	HEMAGGLUTININ (FRAGME	2.08e+02
294	6	1.2	297	14	082071	GLYCOPROTEIN.	2.08e+02	367	6	1.2	344	14	067074	HEMAGGLUTININ (FRAGME	2.08e+02
295	6	1.2	297	14	082070	GLYCOPROTEIN.	2.08e+02	368	6	1.2	344	14	067073	HEMAGGLUTININ (FRAGME	2.08e+02
296	6	1.2	297	14	082067	GLYCOPROTEIN.	2.08e+02	369	6	1.2	344	14	067062	HEMAGGLUTININ (FRAGME	2.08e+02
297	6	1.2	297	14	082066	GLYCOPROTEIN.	2.08e+02	370	6	1.2	344	14	067122	HEMAGGLUTININ (FRAGME	2.08e+02
298	6	1.2	298	14	082068	GLYCOPROTEIN (FRAGMENT	2.08e+02	371	6	1.2	344	14	067122	HEMAGGLUTININ (FRAGME	2.08e+02
299	6	1.2	298	14	001929	MAJOR SURFACE GLYCOPRO	2.08e+02	372	6	1.2	344	14	067081	HEMAGGLUTININ (FRAGME	2.08e+02
300	6	1.2	298	1	058205	298AA LONG HYPOTHETICA	2.08e+02	373	6	1.2	344	14	067059	HEMAGGLUTININ (FRAGME	2.08e+02
301	6	1.2	300	10	022734	SIMILAR TO NAM.	2.08e+02	374	6	1.2	344	14	067059	HEMAGGLUTININ (FRAGME	2.08e+02
302	6	1.2	301	9	021879	MAJOR STRUCTURAL PROTE	2.08e+02	375	6	1.2	344	14	072153	HEMAGGLUTININ (FRAGME	2.08e+02
303	6	1.2	301	9	080113	PUTATIVE MAJOR STRUCTU	2.08e+02	376	6	1.2	344	14	077769	HEMAGGLUTININ (FRAGME	2.08e+02
304	6	1.2	303	2	P72971	HYPOTHETICAL 33.9 KD P	2.08e+02	377	6	1.2	344	14	077771	HEMAGGLUTININ (FRAGME	2.08e+02
305	6	1.2	303	2	066882	TRANSCRIPTIONAL REGULA	2.08e+02	378	6	1.2	344	14	077772	HEMAGGLUTININ (FRAGME	2.08e+02
306	6	1.2	306	2	045286	ORF 2.	2.08e+02	379	6	1.2	344	14	077773	HEMAGGLUTININ (FRAGME	2.08e+02
307	6	1.2	306	10	023551	HYPOTHETICAL 33.5 KD P	2.08e+02	380	6	1.2	344	14	077772	HEMAGGLUTININ (FRAGME	2.08e+02
308	6	1.2	309	1	029477	CONSERVED HYPOTHETICAL	2.08e+02	381	6	1.2	344	14	077768	HEMAGGLUTININ (FRAGME	2.08e+02
309	6	1.2	310	5	001301	F25H9.2 PROTEIN.	2.08e+02	382	6	1.2	344	14	077778	HEMAGGLUTININ (FRAGME	2.08e+02
310	6	1.2	310	10	065269	F6N23.28 PROTEIN.	2.08e+02	383	6	1.2	344	14	077777	HEMAGGLUTININ (FRAGME	2.08e+02
311	6	1.2	311	2	068976	HYPOTHETICAL 35.1 KD P	2.08e+02	384	6	1.2	344	14	077752	HEMAGGLUTININ (FRAGME	2.08e+02
312	6	1.2	311	2	068976	HYPOTHETICAL 35.1 KD P	2.08e+02	385	6	1.2	344	14	077779	HEMAGGLUTININ (FRAGME	2.08e+02

94	7	1.4	869	5	001262	F49C12.15 PROTEIN.	3.55e+00	167	6	1.2	133	10	041863	TRANSPSABLE ELEMENT M	2.08e+02
95	7	1.4	875	5	Q18372	C33D9.1 PROTEIN.	3.55e+00	168	6	1.2	135	3	013568	YPR130CP.	2.08e+02
96	7	1.4	884	4	000302	ERPROR 213-21.	3.55e+00	169	6	1.2	141	2	084321	L11 RIBOSOMAL PROTEIN.	2.08e+02
97	7	1.4	920	4	Q14673	KIAA0164 PROTEIN.	3.55e+00	170	6	1.2	144	2	025757	ATP SYNTHASE F0, SUBUN	2.08e+02
98	7	1.4	932	5	Q17753	SIMILAR TO KINESIN-LIK	3.55e+00	171	6	1.2	145	9	038222	ORF.	2.08e+02
99	7	1.4	952	5	Q94889	PROTEIN KINASE.	3.55e+00	172	6	1.2	145	5	Q17681	COSMID C05H8.	2.08e+02
100	7	1.4	967	3	Q74561	PUTATIVE B-ZIP TRANSCR	3.55e+00	173	6	1.2	148	2	083123	PTS SYSTEM, NITROGEN R	2.08e+02
101	7	1.4	1032	11	Q62780	RNA HELICASE.	3.55e+00	174	6	1.2	151	14	010614	ACMPV ORF1 HOMOLOG.	2.08e+02
102	7	1.4	1044	10	Q80988	PUTATIVE GLYCINE DEHYD	3.55e+00	175	6	1.2	153	2	032455	ORF 5.	2.08e+02
103	7	1.4	1048	11	Q63627	CTD-BINDING SR-LIKE PR	3.55e+00	176	6	1.2	155	2	024931	CONSERVED HYPOTHETICAL	2.08e+02
104	7	1.4	1072	5	Q45232	CO1C7.1 PROTEIN.	3.55e+00	177	6	1.2	155	5	Q21739	RO5D11.7 PROTEIN.	2.08e+02
105	7	1.4	1125	3	Q07660	CHROMOSOME IV READING	3.55e+00	178	6	1.2	161	10	P93490	CELL WALL INVERTASE II	2.08e+02
106	7	1.4	1150	3	Q02511	HYPOTHETICAL 133.0 KD	3.55e+00	179	6	1.2	163	5	Q17155	ALPHA CHAIN OF THE TET	2.08e+02
107	7	1.4	1151	13	P87343	MEMBRANE GUANYLYL CYCL	3.55e+00	180	6	1.2	163	14	Q72789	ENVELOPE PROTEIN (FRAG	2.08e+02
108	7	1.4	1173	14	Q98177	MC006L.	3.55e+00	181	6	1.2	163	5	Q94442	TENTANS ORF'S (A-E) FO	2.08e+02
109	7	1.4	1173	11	Q63624	CTD-BINDING SR-LIKE PR	3.55e+00	182	6	1.2	163	14	Q72791	ENVELOPE PROTEIN (FRAG	2.08e+02
110	7	1.4	1175	14	Q85285	ORF17.	3.55e+00	183	6	1.2	163	4	Q15410	NUCLEIC ACID BINDING P	2.08e+02
111	7	1.4	1179	4	Q14120	DBP-5 NUCLEAR PROTEIN.	3.55e+00	184	6	1.2	163	14	Q72772	ENVELOPE PROTEIN (FRAG	2.08e+02
112	7	1.4	1246	3	Q12276	CHROMOSOME XV READING	3.55e+00	185	6	1.2	163	14	Q72788	ENVELOPE PROTEIN (FRAG	2.08e+02
113	7	1.4	1257	10	Q64768	PUTATIVE SER/THR PROTE	3.55e+00	186	6	1.2	163	14	Q72790	ENVELOPE PROTEIN (FRAG	2.08e+02
114	7	1.4	1268	11	Q63623	CTD-BINDING SR-LIKE PR	3.55e+00	187	6	1.2	163	14	Q72773	ENVELOPE PROTEIN (FRAG	2.08e+02
115	7	1.4	1280	4	Q75158	KIAA0670 PROTEIN (FRAG	3.55e+00	188	6	1.2	166	8	Q79557	NADH DEHYDROGENASE SUB	2.08e+02
116	7	1.4	1288	4	Q15038	KIAA0324 (FRAGMENT).	3.55e+00	189	6	1.2	171	2	Q53838	HYPOTHETICAL PROTEIN (2.08e+02
117	7	1.4	1308	4	Q60307	KIAA0561 PROTEIN (FRAG	3.55e+00	190	6	1.2	172	14	Q69177	GLYCOPROTEIN M (GM) (F	2.08e+02
118	7	1.4	1471	5	Q18245	C27B7.7 PROTEIN.	3.55e+00	191	6	1.2	172	14	P87513	POLYMERASE NSP4 (FRAGM	2.08e+02
119	7	1.4	1590	5	P91152	SIMILAR TO DROSOPHILA	3.55e+00	192	6	1.2	173	1	Q26725	HYPOTHETICAL 19.6 KD P	2.08e+02
120	7	1.4	1791	4	Q60382	KIAA0324 (FRAGMENT).	3.55e+00	193	6	1.2	176	4	Q15380	Y-CHROMOSOME RNA RECOG	2.08e+02
121	7	1.4	1926	11	Q61464	NUCLEAR PROTEIN, NP220	3.55e+00	194	6	1.2	178	2	Q31285	PUTATIVE INNERMEMBRANE	2.08e+02
122	7	1.4	1935	3	Q60072	RNA HELICASE.	3.55e+00	195	6	1.2	178	5	Q62303	K11D2.1 PROTEIN.	2.08e+02
123	7	1.4	1978	4	Q14966	NUCLEAR PROTEIN, NP220	3.55e+00	196	6	1.2	179	14	Q67328	INFLUENZA A/PR/8/34, H	2.08e+02
124	7	1.4	2161	4	Q01668	NEURONAL L-TYPE CALCIU	3.55e+00	197	6	1.2	180	10	Q64541	YUP8H12R.28.	2.08e+02
125	7	1.4	2170	4	Q13931	VOLTAGE-DEPENDENT CALC	3.55e+00	198	6	1.2	182	6	Q77623	CASPASE-3 (FRAGMENT).	2.08e+02
126	7	1.4	2181	4	Q13916	NEUROENDOCRINE/BETA-CE	3.55e+00	199	6	1.2	183	14	Q67650	TYPE III P20 PROTEIN,	2.08e+02
127	7	1.4	2783	5	P91255	FL1F3.2 PROTEIN.	3.55e+00	200	6	1.2	183	11	Q70307	TLS-ASSOCIATED PROTEIN	2.08e+02
128	7	1.4	4128	11	Q81877	DNA-DEPENDENT PROTEIN	3.55e+00	201	6	1.2	183	4	Q05617	TLS-ASSOCIATED PROTEIN	2.08e+02
129	7	1.4	4128	11	P97313	PAIRED BOX PROTEIN PAX	2.08e+02	202	6	1.2	183	2	005817	HYPOTHETICAL 19.5 KD P	2.08e+02
130	6	1.2	39	4	Q92700	RETINOBLASTOMA SUSCEPT	2.08e+02	203	6	1.2	184	5	Q16158	CALCIUM-BINDING PROTEI	2.08e+02
131	6	1.2	41	4	Q92701	RETINOBLASTOMA SUSPECT	2.08e+02	204	6	1.2	184	11	P97297	PREPROADRENOMEDULLIN P	2.08e+02
132	6	1.2	42	14	Q10425	HEMAGGLUTININ (FRAGMEN	2.08e+02	205	6	1.2	184	11	P97453	ADRENOMEDULLIN.	2.08e+02
133	6	1.2	44	8	Q32935	ORF44A.	2.08e+02	206	6	1.2	186	2	P73747	HYPOTHETICAL 21.2 KD P	2.08e+02
134	6	1.2	49	13	Q13031	PROTAMINE 2.	2.08e+02	207	6	1.2	188	3	P87042	CLASS I CHITIN SYNTHAS	2.08e+02
135	6	1.2	52	14	Q92312	ATTACHMENT GLYCOPROTEI	2.08e+02	208	6	1.2	188	2	Q07361	HESA.	2.08e+02
136	6	1.2	70	4	Q14070	(CLONE E35) GENE FROM	2.08e+02	209	6	1.2	189	3	P87044	CLASS II CHITIN SYNTHA	2.08e+02
137	6	1.2	73	8	Q33001	ORF73B.	2.08e+02	210	6	1.2	189	4	Q07823	MAC30 PROTEIN (FRAGMEN	2.08e+02
138	6	1.2	77	1	Q26574	GYRASE A (FRAGMENT).	2.08e+02	211	6	1.2	190	2	Q52136	ESCU.	2.08e+02
139	6	1.2	78	2	Q86357	CONSERVED PROTEIN.	2.08e+02	212	6	1.2	195	8	Q34670	CYTCHROME B (FRAGMENT	2.08e+02
140	6	1.2	80	2	Q69048	RYRMB2.	2.08e+02	213	6	1.2	196	2	Q86675	HYPOTHETICAL 21.4 KD P	2.08e+02
141	6	1.2	81	1	Q54608	GYRASE A (FRAGMENT).	2.08e+02	214	6	1.2	197	11	Q87875	HYPOTHETICAL 21.4 KD P	2.08e+02
142	6	1.2	87	2	Q32324	SIMILAR TO BACILLUS SU	2.08e+02	215	6	1.2	199	14	Q65970	PUTATIVE KERATIN-ASSOC	2.08e+02
143	6	1.2	90	5	Q16596	D2062.6 PROTEIN.	2.08e+02	216	6	1.2	207	5	Q76587	VIRUS COAT PROTEIN.	2.08e+02
144	6	1.2	95	14	Q05535	PROTEIN IR3.	2.08e+02	217	6	1.2	212	8	Q03357	COAT PROTEIN.	2.08e+02
145	6	1.2	99	4	Q15789	NEUTROPHIL PROTEIN (FR	2.08e+02	218	6	1.2	213	8	Q35425	F16G10.5 PROTEIN.	2.08e+02
146	6	1.2	101	1	Q29102	H+-TRANSPORTING ATP SY	2.08e+02	219	6	1.2	213	14	Q77629	ATP SYNTHASE A CHAIN (2.08e+02
147	6	1.2	101	10	Q23452	SIMILARITY TO RIBOSOMA	2.08e+02	220	6	1.2	214	14	Q11784	CYTCHROME B (FRAGMENT	2.08e+02
148	6	1.2	108	2	Q54032	NITROGENASE (FRAGMENT)	2.08e+02	221	6	1.2	214	14	Q86926	ENVELOPE GLYCOPROTEIN	2.08e+02
149	6	1.2	114	8	Q33757	(CJS193) MITOCHONDRIAL	2.08e+02	222	6	1.2	214	3	Q43039	VIF PROTEIN.	2.08e+02
150	6	1.2	114	8	Q33767	(CJS566) MITOCHONDRIAL	2.08e+02	223	6	1.2	216	4	Q43884	TRANSMEMBRANE PROTEIN	2.08e+02
151	6	1.2	115	2	Q56549	ALPHA SUBUNIT OF DINIT	2.08e+02	224	6	1.2	216	4	Q89753	HYPOTHETICAL 24.7 KD P	2.08e+02
152	6	1.2	117	10	Q40816	S-RNASE (FRAGMENT).	2.08e+02	225	6	1.2	217	2	Q51108	INTERNAL ORF TO 0567.	2.08e+02
153	6	1.2	117	14	Q66038	HOST SHUT OFF VIRION P	2.08e+02	226	6	1.2	220	4	Q15376	PUTATIVE RESPONSE REGU	2.08e+02
154	6	1.2	119	14	Q71360	ENVELOPE GLYCOPROTEIN,	2.08e+02	227	6	1.2	220	2	Q51814	OLFACTORY RECEPTOR (FR	2.08e+02
155	6	1.2	121	6	P79109	G-PROTEIN-COUPLED INWA	2.08e+02	228	6	1.2	223	5	Q94611	EHO, COMPLETE GENOME.	2.08e+02
156	6	1.2	121	5	Q15618	ELONGATION FACTOR 1 BE	2.08e+02	229	6	1.2	223	1	Q26310	TRANSFERRIN BINDING PR	2.08e+02
157	6	1.2	123	10	Q40821	S-RNASE (FRAGMENT).	2.08e+02	230	6	1.2	224	11	Q08440	Y-CHROMOSOME RNA RECOG	2.08e+02
158	6	1.2	124	6	Q95294	NUCLEAR PROTEIN (FRAGM	2.08e+02	231	6	1.2	225	5	Q61816	HYPOTHETICAL 25.1 KD P	2.08e+02
159	6	1.2	124	6	Q46582	CYTCHROME C OXIDASE P	2.08e+02	232	6	1.2	225	4	Q43877	CYCLOPHILIN.	2.08e+02
160	6	1.2	125	13	Q05757	OOCTE MATURATION FACT	2.08e+02	233	6	1.2	227	14	Q11286	DNA-DEPENDENT DNA POLY	2.08e+02
161	6	1.2	125	5	Q05757	HOMEOBOX PROTEIN ABDOM	2.08e+02	234	6	1.2	230	4	Q99561	TRANSCRIPTION FACTOR E	2.08e+02
162	6	1.2	125	5	Q05757	OOCTE MATURATION FACT	2.08e+02	235	6	1.2	231	4	Q14911	B0511.5 PROTEIN.	2.08e+02
163	6	1.2	125	5	Q05757	OOCTE MATURATION FACT	2.08e+02	236	6	1.2	231	4	Q14911	OLFACTORY RECEPTOR (FR	2.08e+02
164	6	1.2	125	5	Q05757	OOCTE MATURATION FACT	2.08e+02	237	6	1.2	231	4	Q14911	HYPOTHETICAL 24.6 KD P	2.08e+02
165	6	1.2	125	5	Q05757	OOCTE MATURATION FACT	2.08e+02	238	6	1.2	231	4	Q14911	NK RECEPTOR.	2.08e+02
166	6	1.2	125	5	Q05757	OOCTE MATURATION FACT	2.08e+02	239	6	1.2	231	2	Q69442	TAX INTERACTION PROTEI	2.08e+02
167	6	1.2	133	10	041863	TRANSPSABLE ELEMENT M	2.08e+02	239	6	1.2	231	2	Q69442	SUCRASE (EC 3.2.1.26)	2.08e+02



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_p protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 10 11:40:06 1999; MasPar time 28.83 Seconds

Tabular output not generated. 952.389 Million cell updates/sec

Title: >US-08-287-669-19
Description: (1-503) from US08287669.pep
Perfect Score: 503
Sequence: 1 MMRQDRSLERINIMFSSH.....MTSRLKRFYFWPEARNSAV 503

Scoring table: TABLE uninterpretable
Gap 60

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database:

sptrembl9
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 3.697; Variance 0.424; scale 8.719

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	416	82.7	495	5	002229	C48D1.2 PROTEIN.	0.00e+00
2	8	1.6	102	8	079747	CYTOSOL B (FRAGMENT	3.37e-02
3	8	1.6	174	3	074357	POSSIBLE SPLICE FACTOR	3.37e-02
4	8	1.6	263	3	013854	SPAC19G12.17C PROTEIN	3.37e-02
5	8	1.6	301	2	007016	HYPOTHETICAL 33.7 KD P	3.37e-02
6	8	1.6	327	5	002155	CODED FOR BY C. ELEGAN	3.37e-02
7	8	1.6	456	3	042840	HYPOTHETICAL 46.7 KD P	3.37e-02
8	8	1.6	619	5	018207	Y48E1C.3 PROTEIN.	3.37e-02
9	8	1.6	1045	5	077435	EG:34F3.4 PROTEIN.	3.37e-02
10	7	1.4	62	6	018749	PROTAMINE P1.	3.55e+00
11	7	1.4	128	14	011288	HYPOTHETICAL 13.2 KD P	3.55e+00
12	7	1.4	130	5	019225	F08H9.2 PROTEIN.	3.55e+00
13	7	1.4	134	2	044177	ORE134.	3.55e+00
14	7	1.4	134	5	020104	THIS GENE MAY BEGIN IN	3.55e+00
15	7	1.4	162	14	065580	UL6 PROTEIN (FRAGMENT)	3.55e+00
16	7	1.4	163	5	001690	COSMID T08B2.	3.55e+00
17	7	1.4	175	13	090768	CD3 GLYCOPROTEIN.	3.55e+00
18	7	1.4	179	5	018409	SIMILAR TO PRE-MRNA SP	3.55e+00
19	7	1.4	185	5	P91177	COSMID C50F2.	3.55e+00
20	7	1.4	186	11	Q60995	PANETH CELL ENHANCED E	3.55e+00

21	7	1.4	197	2	068247	HYPOTHETICAL 22.4 KD P	3.55e+00
22	7	1.4	197	5	Q17626	C04G2.8 PROTEIN.	3.55e+00
23	7	1.4	203	5	P91026	CODED FOR BY C. ELEGAN	3.55e+00
24	7	1.4	203	5	P91497	CODED FOR BY C. ELEGAN	3.55e+00
25	7	1.4	203	8	Q33765	MITOCHONDRIAL DNA, COM	3.55e+00
26	7	1.4	218	2	P95075	FUCA.	3.55e+00
27	7	1.4	235	10	P93396	TRANSFORMER-SR RIBONUC	3.55e+00
28	7	1.4	237	1	080496	T12M4.19 PROTEIN.	3.55e+00
29	7	1.4	253	10	Q27764	PHOSPHATE TRANSPORT SY	3.55e+00
30	7	1.4	268	3	Q99394	YOR3251C FROM CHROMOSO	3.55e+00
31	7	1.4	281	5	Q19208	COSMID F08F8.	3.55e+00
32	7	1.4	286	10	Q82021	PUTATIVE ARGININE/SERI	3.55e+00
33	7	1.4	288	4	Q15815	HTRA2-BETA.	3.55e+00
34	7	1.4	288	11	Q64283	SILICA-INDUCED PROTEIN	3.55e+00
35	7	1.4	299	5	P91125	CODED FOR BY C. ELEGAN	3.55e+00
36	7	1.4	303	11	Q88550	CASPASE-7.	3.55e+00
37	7	1.4	312	5	Q23121	W02B12.3 PROTEIN.	3.55e+00
38	7	1.4	315	14	Q92485	PP34-POLYHERON CALYX-P	3.55e+00
39	7	1.4	322	5	Q23796	HNRNP PROTEIN.	3.55e+00
40	7	1.4	332	11	Q35986	ZIS.	3.55e+00
41	7	1.4	337	13	Q91708	C4SR PROTEIN.	3.55e+00
42	7	1.4	346	2	Q86705	HYPOTHETICAL 38.6 KD P	3.55e+00
43	7	1.4	352	5	Q19209	COSMID F08F8.	3.55e+00
44	7	1.4	352	5	Q20393	F4AD12.4 PROTEIN.	3.55e+00
45	7	1.4	361	14	Q65558	UL31.	3.55e+00
46	7	1.4	378	11	Q60854	SERINE PROTEASE INHIBI	3.55e+00
47	7	1.4	379	14	Q67475	40 KD PROTEIN.	3.55e+00
48	7	1.4	386	5	Q61646	SPLICING FACTOR SRP54.	3.55e+00
49	7	1.4	410	10	Q41018	GLOBULIN-2.	3.55e+00
50	7	1.4	429	3	Q13897	HYPOTHETICAL 48.2 KD P	3.55e+00
51	7	1.4	429	3	Q67776	HYPOTHETICAL 47.8 KD P	3.55e+00
52	7	1.4	431	2	Q66238	ATP BINDING COMPONENT	3.55e+00
53	7	1.4	431	2	Q70098	ATP BINDING COMPONENT	3.55e+00
54	7	1.4	431	2	Q66242	ATP BINDING COMPONENT	3.55e+00
55	7	1.4	431	2	Q47591	PUTATIVE ATP BINDING C	3.55e+00
56	7	1.4	435	5	Q26692	TBRM1.	3.55e+00
57	7	1.4	438	1	Q27239	PHOSPHONACETALDEHYDE	3.55e+00
58	7	1.4	440	2	Q66638	HYPOTHETICAL 50.3 KD P	3.55e+00
59	7	1.4	456	3	Q06340	CHROMOSOME IV COSMID 9	3.55e+00
60	7	1.4	460	3	Q02321	EXOCYTOBIOLHYDROLASE.	3.55e+00
61	7	1.4	467	5	Q76777	UDP-N-ACETYLGLUCOSAMIN	3.55e+00
62	7	1.4	469	14	Q56937	REGULATORY PROTEIN E2.	3.55e+00
63	7	1.4	470	5	Q22919	COSMID C37C3.	3.55e+00
64	7	1.4	471	2	Q67995	ZINC-PROTEASE TRANSPOR	3.55e+00
65	7	1.4	478	14	Q56943	REGULATORY PROTEIN E2.	3.55e+00
66	7	1.4	484	4	Q05519	ARGININE-RICH 54 KD NU	3.55e+00
67	7	1.4	488	10	Q41017	GLOBULIN-1.	3.55e+00
68	7	1.4	488	10	Q04528	SEQUENCE OF BAC F20P5	3.55e+00
69	7	1.4	503	8	Q37612	CYTOSOL C OXIDASE S	3.55e+00
70	7	1.4	505	5	Q24898	NUCLEAR PROTEIN.	3.55e+00
71	7	1.4	524	4	Q14498	SPLICING FACTOR.	3.55e+00
72	7	1.4	530	4	Q14499	SPLICING FACTOR.	3.55e+00
73	7	1.4	534	5	Q18515	NUCLEOPORIN P62.	3.55e+00
74	7	1.4	535	3	Q03359	D8035.17P.	3.55e+00
75	7	1.4	558	5	Q18254	Y57G11C.9 PROTEIN.	3.55e+00
76	7	1.4	582	10	Q82649	MAP3K ALPHA PROTEIN KI	3.55e+00
77	7	1.4	588	2	Q69007	MALTOGENIC AMYLASE.	3.55e+00
78	7	1.4	591	10	Q82667	MAP3K ALPHA 1 PROTEIN	3.55e+00
79	7	1.4	603	10	Q49485	PHOSPHOGLYCERATE DEHYD	3.55e+00
80	7	1.4	620	11	Q08984	RAT NBP60.	3.55e+00
81	7	1.4	656	2	Q55064	RUB.	3.55e+00
82	7	1.4	656	2	Q55710	THIG PROTEIN.	3.55e+00
83	7	1.4	672	5	Q19854	F28C1.1 PROTEIN.	3.55e+00
84	7	1.4	682	5	Q44553	K06A5.1 PROTEIN.	3.55e+00
85	7	1.4	688	14	Q65575	PUTATIVE VIRION PROTEI	3.55e+00
86	7	1.4	690	5	Q18413	SIMILAR TO KINESIN-LIK	3.55e+00
87	7	1.4	702	5	Q20564	HYPOTHETICAL 78.8 KD P	3.55e+00
88	7	1.4	717	5	Q10954	RGA AND ATU GENES, COM	3.55e+00
89	7	1.4	724	5	Q94546	F32B4.4 PROTEIN.	3.55e+00
90	7	1.4	771	5	Q62203	MAJOR CAPSID PROTEIN P	3.55e+00
91	7	1.4	783	14	Q67723	U5 SNRNP 100 KD PROTEI	3.55e+00
92	7	1.4	820	4	Q43188	F32H2.3 PROTEIN.	3.55e+00
93	7	1.4	824	5	P91870		3.55e+00


```
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94366758.
RA WONG W.T., CARLOMAGNO F., DRUCK T., BARLETTA C., CROCE C.M.,
RA HUEBNER K., KRAUS M.H., DI FIORE P.P.;
RT "Evolutionary conservation of the EPS8 gene and its mapping to human
RT chromosome 12q23-q24.";
RL ONCOGENE 9:3057-3061(1994).
CC -1- FUNCTION: UPON BINDING TO EGF RECEPTOR ENHANCES EGF-DEPENDENT
CC MITOGENIC SIGNALS. CAN BIND MULTIPLE CELLULAR TARGETS.
CC -1- PTM: PHOSPHORYLATED BY SEVERAL RECEPTOR TYROSINE KINASES.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES ANALYZED, INCLUDING
CC HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY AND
CC PANCREAS. EXPRESSED IN ALL EPITHELIAL AND FIBROBLASTIC LINES
CC EXAMINED AND IN SOME, BUT NOT ALL, HEMATOPOIETIC CELLS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SPLIT PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U12535; G530823; -.
DR MIM: 600206; -.
DR PROSITE: PS50002; SH3; 1.
DR PFAM: PF00018; SH3; 1.
DR HSSP: Q08509; 1A0J.
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 69 129 PH (FIRST PART).
FT DOMAIN 210 213 POLY-PRO.
FT DOMAIN 322 325 POLY-PRO.
FT DOMAIN 381 414 PH (SECOND PART).
FT DOMAIN 421 440 PRO-RICH.
FT DOMAIN 532 591 SH3.
FT DOMAIN 615 651 PRO-RICH.
FT DOMAIN 659 664 POLY-SER.
SQ SEQUENCE 822 AA; 91881 MW; 501CF5AE CRC32;

Query Match 1.6%; Score 8; DB 1; Length 822;
Best Local Similarity 100.0%; Pred. No. 2.00e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

53 RDSVSSVS 60
11111111
QY 130 RDSVSSVS 137

RESULT 15
ID EML_ENSMI STANDARD; PRT; 34 AA.
AC P27205;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE SPERM PROTEIN EMI (FRAGMENT).
OS ENSIS MINOR (RAZOR SHELL) (MINOR JACKKNIFE CLAM).
OC EUKARYOTA; METAZOA; MOLUSCA; BIVALVIA; HETEROCONCHIA; VENEROIDA;
OC SOLENIDAE; ENSIS.
RN [1]
RP SEQUENCE.
RC TISSUE=SPERM;
RX MEDLINE: 92190249.
RA GIANCOTTI V., BURATTI E., SANTUCCI A., NERI P., CRANE-ROBINSON C.;
RT "Molluscan sperm proteins: Ensis minor.";
RL BIOCHIM. BIOPHYS. ACTA 1119:296-302(1992).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: SPERM.
CC -1- SIMILARITY: PROTAMINE-LIKE.
DR PIR: S21080; S21080.
KW CHROMOSOMAL PROTEIN; NUCLEAR PROTEIN; DNA-BINDING; SPERM; REPEAT.
```

```
FT DOMAIN 3 16 7 X 2 AA TANDEM REPEATS OF S-[KR].
FT NON_TER 34 34
SQ SEQUENCE 34 AA; 3531 MW; 77A8D51E CRC32;

Query Match 1.4%; Score 7; DB 1; Length 34;
Best Local Similarity 100.0%; Pred. No. 2.20e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6 RSRSRSR 12
11111111
QY 150 RSRSRSR 156

Search completed: Tue Aug 10 11:39:50 1999
Job time : 77 secs.
```


DR EMBL: D90913; G1653440; -
DR PROSITE: PS00933; FG_Y_KINASES_1; FALSE_NEG.
DR PROSITE: PS00445; FG_Y_KINASES_2; 1.
DR PFAM: PF00370; FG_Y; 1.
DR HSSP: P08859; 1GLB.
KW GLYCEROL METABOLISM; TRANSFERASE; KINASE.
SQ SEQUENCE 495 AA; 54351 MW; C7247592 CRC32;

Query Match 1.6%; Score 8; DB 1; Length 495;
Best Local Similarity 100.0%; Pred. No. 2.00e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 338 LARSVDN 345
|||
QY 91 LARSVDN 98

RESULT 12
ID VE2_HP47 STANDARD; PRT; 506 AA.
AC P22420;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE REGULATORY PROTEIN E2.
GN E2.
OS HUMAN PAPILLOMAVIRUS TYPE 47.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAVOVIRIDAE; PAPILLOMAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90281611.
KY Kiyono T., Adachi A., Ishibashi M.;
RT "Genome organization and taxonomic position of human papillomavirus
type 47 inferred from its DNA sequence.";
RL VIROLOGY 177:401-405(1990).

CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACNNNNNGGT-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SUBUNIT: BINDS DNA AS A DIMER.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M32305; G333067; -
DR PIR: D35324; W2WL47.
DR PFAM: PF00508; E2_N; 1.
DR PFAM: PF00511; E2_C; 1.
DR HSSP: P03122; 2BOP.
KW EARLY PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING;
KW TRANS-ACTING FACTOR; DNA REPLICATION; REPRESSOR; NUCLEAR PROTEIN.
SQ SEQUENCE 506 AA; 57478 MW; F93097EF CRC32;

Query Match 1.6%; Score 8; DB 1; Length 506;
Best Local Similarity 100.0%; Pred. No. 2.00e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 302 RARSRSRS 309
|||||
QY 148 RARSRSRS 155

RESULT 13
ID EPS8_MOUSE STANDARD; PRT; 821 AA.

AC Q08509;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8.
GN EPS8.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94008987.
RA FAZIOLE F., MINICHIELLO L., MATOSKA V., CASTAGNINO P., MIKI T.,
RA WONG W.T., DI FIORE P.P.;
RT "Eps8, a substrate for the epidermal growth factor receptor kinase,
RT enhances EGF-dependent mitogenic signals.";
RL EMBO J. 12:3799-3808(1993).

CC [2]
CC X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 532-591.
CC MEDLINE: 97448677.
CC KISHAN K.V.R., SCITA G., WONG W.T., DI FIORE P.P., NEWCOMER M.E.;
CC "The SH3 domain of Eps8 exists as a novel intertwined dimer.";
CC NAT. STRUCT. BIOL. 4:739-743(1997).
CC -1- FUNCTION: UPON BINDING TO EGF RECEPTOR ENHANCES EGF-DEPENDENT
CC MITOGENIC SIGNALS. CAN BIND MULTIPLE CELLULAR TARGETS.
CC -1- PTM: PHOSPHORYLATED BY SEVERAL RECEPTOR TYROSINE KINASES.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SPLIT PH DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L21671; G309217; -
DR PDB: LAOJ; 08-JUL-98.
DR MGD: MGI:104684; EPS8.
DR PROSITE: PS50002; SH3; 1.
DR PFAM: PF00018; SH3; 1.
KW SH3 DOMAIN; PHOSPHORYLATION; 3D-STRUCTURE.
FT DOMAIN 69 129 PH (FIRST PART).
FT DOMAIN 210 213 POLY-PRO.
FT DOMAIN 322 325 POLY-PRO.
FT DOMAIN 381 414 PH (SECOND PART).
FT DOMAIN 421 440 PRO-RICH.
FT DOMAIN 532 591 SH3.
FT DOMAIN 620 650 PRO-RICH.
FT DOMAIN 658 663 POLY-SER.
SQ SEQUENCE 821 AA; 91738 MW; 56D28642 CRC32;

Query Match 1.6%; Score 8; DB 1; Length 821;
Best Local Similarity 100.0%; Pred. No. 2.00e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 53 RDSVSVS 60
|||||
QY 130 RDSVSVS 137

RESULT 14
ID EPS8_HUMAN STANDARD; PRT; 822 AA.
AC Q12929;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8.
GN EPS8.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.

CC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
CC PLERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYROIDEA;
CC DROSOPHILIDAE; DROSOPHILA.
CC [1]
CC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
CC TISSUE-EMBRYO;
CC MEDLINE; 96413646.
CC RX RUDNER D.Z., KANAR R., BREGER K.S., RIO D.C.;
CC "Mutations in the small subunit of the Drosophila U2AF splicing
CC factor cause lethality and developmental defects."
CC RT PROC. NATL. ACAD. SCI. U.S.A. 93:10333-10337(1996).
CC RL -1- FUNCTION: NECESSARY FOR THE SPLICING OF PRE-MRNA. BINDS TO THE
CC POLYPYRIMIDINE TRACT OF INTRONS EARLY DURING SPLICEOSOME ASSEMBLY
CC (BY SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH A 65 KD PROTEIN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67066; G1621615; -
CC DR FLYBASE; FBgn0017457; U2af38.
CC DR PFAM; PF00076; rtm; 1.
CC DR PFAM; PF00642; zf-CCCH; 2.
CC KM NUCLEAR PROTEIN; RNA-BINDING; MRNA SPLICING.
CC FT DOMAIN 46 174 INVOLVED IN BINDING U2AF-50
CC (BY SIMILARITY).
CC FT DOMAIN 180 213 ARG/SER-RICH (RS DOMAIN).
CC FT DOMAIN 190 197 POLY-ARG.
CC FT DOMAIN 252 262 POLY-GLY.
CC SQ SEQUENCE 264 AA; 29855 MW; C267A411 CRC32;
CC -----
CC Query Match 1.6%; Score 8; DB 1; Length 264;
CC Best Local Similarity 100.0%; Pred. No. 2.00e-02;
CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC Db 193 RARSRSRS 200
CC QY 148 RARSRSRS 155
CC [1]
CC RESULT 10
CC ID ICE3_XENLA STANDARD; PRT; 282 AA.
CC AC P55866;
CC DT 01-NOV-1997 (REL. 35, CREATED)
CC DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
CC DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
CC DE APOPAIN PRECURSOR (EC 3.4.22.-) (CYSTEINE PROTEASE CPP32) (CASPASE-3)
CC (CASP-3) (XCPP32).
CC GN CASP3.
CC OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
CC OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
CC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE; 97184166.
CC RA YAOITA Y., NAKAJIMA K.;
CC RT "Induction of apoptosis and CPP32 expression by thyroid hormone in a
CC myoblastic cell line derived from tadpole tail."
CC RL J. BIOL. CHEM. 272:5122-5127(1997).
CC -1- FUNCTION: IMPORTANT MEDIATOR OF APOPTOSIS. AT THE ONSET OF
CC APOPTOSIS IT PROTEOLYTICALLY CLEAVES POLY(ADP-RIBOSE) POLYMERASE
CC (PARP) AT A 216-ASP-1-GLY-217 BOND (BY SIMILARITY).
CC -1- SUBUNIT: HETERODIMER OF A 17 KD (P17) AND A 12 KD (P12) SUBUNITS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- THE SUBUNITS ARE DERIVED FROM THE PRECURSOR SEQUENCE BY A PROBABLE

CC AUTOCATALYTIC MECHANISM AND PROBABLY BY OTHER CASPASES (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CC CASPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D89784; D1014718; -
CC DR PROSITE; PS01121; CASPASE_HIS; 1.
CC DR PROSITE; PS01122; CASPASE_CYS; 1.
CC DR PFAM; PF00655; ICE_P10; 1.
CC DR PFAM; PF00656; ICE_P20; 1.
CC DR HSSP; P42574; 1PAU.
CC KM HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS.
CC FT PROPEP 1 ?
CC FT CHAIN 186 ?
CC FT CHAIN 187 282 APOPAIN P17 SUBUNIT.
CC FT ACT_SITE 131 131 APOPAIN P12 SUBUNIT.
CC FT ACT_SITE 174 174 BY SIMILARITY.
CC FT ACT_SITE 174 174 BY SIMILARITY.
CC SQ SEQUENCE 282 AA; 32124 MW; E9608701 CRC32;
CC -----
CC Query Match 1.6%; Score 8; DB 1; Length 282;
CC Best Local Similarity 100.0%; Pred. No. 2.00e-02;
CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC Db 56 GMCLIIINN 63
CC QY 243 GMCLIIINN 250
CC [1]
CC RESULT 11
CC ID GLPK_SYNY3 STANDARD; PRT; 495 AA.
CC AC P74260;
CC DT 01-NOV-1997 (REL. 35, CREATED)
CC DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
CC DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
CC DE GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL 3-PHOSPHOTRANSFERASE)
CC (GLYCEROKINASE) (GK).
CC DE GLPK OR SLR1672.
CC GN GLPK OR SLR1672.
CC OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
CC OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE; 97061201.
CC RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
CC MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
CC HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K.,
CC OKUMURA S., SHIMO S., TAKEUCHI C., WADA T., WATANABE A.,
CC YAMADA M., YASUDA M., TABATA S.;
CC RT "Sequence analysis of the genome of the unicellular cyanobacterium
CC Synechocystis sp. strain PCC6803. II. Sequence determination of the
CC entire genome and assignment of potential protein-coding regions."
CC RL DNA RES. 3:109-136(1996).
CC -1- CATALYTIC ACTIVITY: ATP + GLYCEROL = ADP + GLYCEROL 3-PHOSPHATE.
CC -1- PATHWAY: RATE-LIMITING STEP IN GLYCEROL UTILIZATION.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUCONOKINASE /
CC GLYCEROKINASE / XYLULOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

RT "Purification, cloning, and expression of ciliary neurotrophic factor
(CNTF).";
RL SCIENCE 246:1023-1025(1989).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE: 90256829.
RA LIN L.-F.H., ARMES L.G., SOMMER A., SMITH D.J., COLLINS F.;
RT "Isolation and characterization of ciliary neurotrophic factor from
rabbit sciatic nerves";
RL J. BIOL. CHEM. 265:8942-8947(1990).
CC -1- FUNCTION: CNTF IS A SURVIVAL FACTOR FOR VARIOUS NEURONAL CELL
CC TYPES. SEEMS TO PREVENT THE DEGENERATION OF MOTOR AXONS AFTER
CC AXOTOMY.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: NERVOUS SYSTEM.
CC -1- SIMILARITY: BELONGS TO THE CNTF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M29828; G164890; -.
DR PIR: A40082; UNRBCF.
DR PFAM: PF01110; CNTF; 1.
DR HSSP: P26441; 1CNT.
KM GROWTH FACTOR; NEURONE.
SQ SEQUENCE 199 AA; 22662 MW; BB9A0976 CRC32;

Db 50 LDSVDGVP 57
QY 370 LDSVDGVP 377

Query Match 1.6%; Score 8; DB 1; Length 199;
Best Local Similarity 100.0%; Pred. No. 2.00e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7
ID CNTF_PIG STANDARD; PRT; 200 AA.
AC 002732;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CILIARY NEUROTROPHIC FACTOR (CNTF).
GN CNTF.
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RA PIEDRAHITA J.A., WEAKS R.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: CNTF IS A SURVIVAL FACTOR FOR VARIOUS NEURONAL CELL
CC TYPES. SEEMS TO PREVENT THE DEGENERATION OF MOTOR AXONS AFTER
CC AXOTOMY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: NERVOUS SYSTEM (BY SIMILARITY).
CC -1- SIMILARITY: HIGH, WITH OTHER MAMMALIAN CNTF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U57644; G2072395; -.
DR PFAM: PF01110; CNTF; 1.
OS

DR HSSP: P26441; 1CNT.
KM GROWTH FACTOR; NEURONE.
SQ SEQUENCE 200 AA; 22718 MW; CDFEB09 CRC32;

Db 50 LDSVDGVP 57
QY 370 LDSVDGVP 377

Query Match 1.6%; Score 8; DB 1; Length 200;
Best Local Similarity 100.0%; Pred. No. 2.00e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 8
ID CNTF_RAT STANDARD; PRT; 200 AA.
AC P20294;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE CILIARY NEUROTROPHIC FACTOR (CNTF).
GN CNTF.
OS RATRUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATRUS.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE: 90081871.
RA STOECKLI K.A., LOTSPEICH F., SENDTNER M., MASIAKOWSKI P., CARROLL P.,
RA GOETZ R., LINDHOLM D., THOENEN H.;
RT "Molecular cloning, expression and regional distribution of rat
RT ciliary neurotrophic factor";
RL NATURE 342:920-923(1989).
CC -1- FUNCTION: CNTF IS A SURVIVAL FACTOR FOR VARIOUS NEURONAL CELL
CC TYPES. SEEMS TO PREVENT THE DEGENERATION OF MOTOR AXONS AFTER
CC AXOTOMY.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: NERVOUS SYSTEM.
CC -1- SIMILARITY: BELONGS TO THE CNTF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17457; G55969; -.
DR PIR: S08144; UNRTCF.
DR PFAM: PF01110; CNTF; 1.
DR HSSP: P26441; 1CNT.
KM GROWTH FACTOR; NEURONE.
SQ SEQUENCE 200 AA; 22854 MW; A1693EF9 CRC32;

Db 50 LDSVDGVP 57
QY 370 LDSVDGVP 377

Query Match 1.6%; Score 8; DB 1; Length 200;
Best Local Similarity 100.0%; Pred. No. 2.00e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 9
ID U2AG_DROME STANDARD; PRT; 264 AA.
AC Q94535;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE SPLICING FACTOR U2AF 38 KD SUBUNIT (U2 AUXILIARY FACTOR 38 KD
DE SUBUNIT) (U2 SNRNP AUXILIARY FACTOR SMALL SUBUNIT).
GN U2AF38.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).

DR	PFAM; PF00655; ICE_p10; 1.	DR	PFAM; PF00656; ICE_p20; 1.	DR	HSSP; P42574; 1PAU.	KM	HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS; ALTERNATIVE SPLICING.	FT	PROPEP	1	23	CASPASE-7 SUBUNIT P20.	FT	CHAIN	24	198	FT	PROPEP	199	206	FT	CHAIN	207	303	CASPASE-7 SUBUNIT P11.	FT	ACT_SITE	144	144	FT	ACT_SITE	186	186	FT	MUTAGEN	186	186	FT	VARSPPLIC	1	1	FT	VARSPPLIC	149	303	FT	VARSPPLIC	149	303	FT	CONFLICT	194	194	SO	SEQUENCE	303 AA;	34276 MM;	G -> A (IN REF. 1).	754CA81F CRC32;
DR	PFAM; PF00655; ICE_p10; 1.	DR	PFAM; PF00656; ICE_p20; 1.	DR	HSSP; P42574; 1PAU.	KM	HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS; ALTERNATIVE SPLICING.	FT	PROPEP	1	23	CASPASE-7 SUBUNIT P20.	FT	CHAIN	24	198	FT	PROPEP	199	206	FT	CHAIN	207	303	CASPASE-7 SUBUNIT P11.	FT	ACT_SITE	144	144	FT	ACT_SITE	186	186	FT	MUTAGEN	186	186	FT	VARSPPLIC	1	1	FT	VARSPPLIC	149	303	FT	CONFLICT	194	194	SO	SEQUENCE	303 AA;	34276 MM;	G -> A (IN REF. 1).	754CA81F CRC32;				
FT	PFAM; PF00655; ICE_p10; 1.	FT	PFAM; PF00656; ICE_p20; 1.	FT	HSSP; P42574; 1PAU.	FT	HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS; ALTERNATIVE SPLICING.	FT	PROPEP	1	23	CASPASE-7 SUBUNIT P20.	FT	CHAIN	24	198	FT	PROPEP	199	206	FT	CHAIN	207	303	CASPASE-7 SUBUNIT P11.	FT	ACT_SITE	144	144	FT	ACT_SITE	186	186	FT	MUTAGEN	186	186	FT	VARSPPLIC	1	1	FT	VARSPPLIC	149	303	FT	CONFLICT	194	194	SO	SEQUENCE	303 AA;	34276 MM;	G -> A (IN REF. 1).	754CA81F CRC32;				
FT	PFAM; PF00655; ICE_p10; 1.	FT	PFAM; PF00656; ICE_p20; 1.	FT	HSSP; P42574; 1PAU.	FT	HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS; ALTERNATIVE SPLICING.	FT	PROPEP	1	23	CASPASE-7 SUBUNIT P20.	FT	CHAIN	24	198	FT	PROPEP	199	206	FT	CHAIN	207	303	CASPASE-7 SUBUNIT P11.	FT	ACT_SITE	144	144	FT	ACT_SITE	186	186	FT	MUTAGEN	186	186	FT	VARSPPLIC	1	1	FT	VARSPPLIC	149	303	FT	CONFLICT	194	194	SO	SEQUENCE	303 AA;	34276 MM;	G -> A (IN REF. 1).	754CA81F CRC32;				
FT	PFAM; PF00655; ICE_p10; 1.	FT	PFAM; PF00656; ICE_p20; 1.	FT	HSSP; P42574; 1PAU.	FT	HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS; ALTERNATIVE SPLICING.	FT	PROPEP	1	23	CASPASE-7 SUBUNIT P20.	FT	CHAIN	24	198	FT	PROPEP	199	206	FT	CHAIN	207	303	CASPASE-7 SUBUNIT P11.	FT	ACT_SITE	144	144	FT	ACT_SITE	186	186	FT	MUTAGEN	186	186	FT	VARSPPLIC	1	1	FT	VARSPPLIC	149	303	FT	CONFLICT	194	194	SO	SEQUENCE	303 AA;	34276 MM;	G -> A (IN REF. 1).	754CA81F CRC32;				
FT	PFAM; PF00655; ICE_p10; 1.	FT	PFAM; PF00656; ICE_p20; 1.	FT	HSSP; P42574; 1PAU.	FT	HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS; ALTERNATIVE SPLICING.	FT	PROPEP	1	23	CASPASE-7 SUBUNIT P20.	FT	CHAIN	24	198	FT	PROPEP	199	206	FT	CHAIN	207	303	CASPASE-7 SUBUNIT P11.	FT	ACT_SITE	144	144	FT	ACT_SITE	186	186	FT	MUTAGEN	186	186	FT	VARSPPLIC	1	1	FT	VARSPPLIC	149	303	FT	CONFLICT	194	194	SO	SEQUENCE	303 AA;	34276 MM;	G -> A (IN REF. 1).	754CA81F CRC32;				
FT	PFAM; PF00655; ICE_p10; 1.	FT	PFAM; PF00656; ICE_p20; 1.	FT	HSSP; P42574; 1PAU.	FT	HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS; ALTERNATIVE SPLICING.	FT	PROPEP	1	23	CASPASE-7 SUBUNIT P20.	FT	CHAIN	24	198	FT	PROPEP	199	206	FT	CHAIN	207	303	CASPASE-7 SUBUNIT P11.	FT	ACT_SITE	144	144	FT	ACT_SITE	186	186	FT	MUTAGEN	186	186	FT	VARSPPLIC	1	1	FT	VARSPPLIC	149	303	FT	CONFLICT	194	194	SO	SEQUENCE	303 AA;	34276 MM;	G -> A (IN REF. 1).	754CA81F CRC32;				
FT	PFAM; PF00655; ICE_p10; 1.	FT	PFAM; PF00656; ICE_p20; 1.	FT	HSSP; P42574; 1PAU.	FT	HYDROLASE; THIOL PROTEASE; ZYMOGEN; APOPTOSIS; ALTERNATIVE SPLICING.	FT	PROPEP	1	23	CASPASE-7 SUBUNIT P20.	FT	CHAIN	24	198	FT	PROPEP	199	206	FT	CHAIN	207	303	CASPASE-7 SUBUNIT P11.	FT	ACT_SITE	144	144	FT	ACT_SITE	186	186	FT	MUTAGEN	186	186	FT	VARSPPLIC	1	1	FT	VARSPPLIC	149	303	FT	CONFLICT	194	194	SO	SEQUENCE	303 AA;	34276 MM;	G -> A (IN REF. 1).	754CA81F CRC32;				
FT	PFAM; PF00655; ICE_p10; 1.	FT																																																									

Query Match	1.8%;	Score 9;	DB 1;	Length 303;
Best Local Similarity	100.0%;	Pred. No. 1.11e-04;		
Matches	9;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

Db	142	LSHGEENV	150
QY	313	LSHGEENV	321

```

      4
RESULT  STANDARD; PRT; 58 AA.
ID      BVCP_GVCL
AC      P41726;
DT      01-NOV-1995 (REL. 32, CREATED)
TM      01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT      15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE      DNA-BINDING PROTEIN (ARGININE RICH PROTEIN 7.3 KD) (BASIC VIRAL CORE
        PROTEIN) (NUCLEOCAPSID PROTEIN).
GN      P7.3.
US      CRYPTOPLHLEBIA LEUCOTRETA GRANULOSIS VIRUS (CLGV) (CRYPTOPHLEBIA
        LEUCOTRETA GRANULOVIRUS).
OC      VIRUSES; DSDNA VIRUSES, NO RNA STAGE; BACULOVIRIDAE; GRANULOVIRUS.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CV3;
PX      MEDLINE; 94292936.
RA      JEHLE J.A., BACKHAUS H.;
RT      "Genome organization of the DNA-binding protein gene region of
RT      Cryptophlebia leucotreta granulosis virus is closely related to that
RT      of nuclear polyhedrosis viruses.";
RL      J. GEN. VIROL. 75:1815-1820(1994).
CC      -I- FUNCTION: THOUGHT TO BE RESPONSIBLE FOR DNA CONDENSATION DURING
        PACKAGING OF THE NUCLEOCAPSIDS.
CC      -I- PTM: PROBABLY PHOSPHORYLATED IN INFECTED CELLS.
CC      -I- SIMILARITY: STRONG TO ARGININE RICH DNA-BINDING PROTEIN OF OTHER
        BACULOVIRUSES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
        or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: X77048; G515334; -
KW      NUCLEOCAPSID; DNA-BINDING; LATE PROTEIN; REPEAT; PHOSPHORYLATION.
SQ      SEQUENCE 58 AA; 7291 MW; 0754FDE0 CRC32;

Query Match          1.6%; Score 8; DB 1; Length 58;
Best Local Similarity 100.0%; Pred. No. 2.00e-02;

```

Matches	8;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Db	37	RSRSRSRA	44						
QY	150	RSRSRSRA	157						

```
Db      37  RSRSRRA 44
      |||||
QY     150 RSRSRRA 157
```

```

RESULT      5
ID          CNTF_MOUSE      STANDARD;      PRT;      198 AA.
AC          P51642;
DF          01-OCT-1996 (REL. 34, CREATED)
DT          01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT          01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE          CILIARY NEUROTROPHIC FACTOR (CNTF).
CC          CNTF.
OS          MUS MUSCULUS (MOUSE).
OC          EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC          RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN          [1]
RP          SEQUENCE FROM N.A.
RC          TISSUE=TESTIS;
RX          MEDLINE; 95137394.
RA          SAOTOME Y., WINTER C.G., HIRSH D.;
RT          "A widely expressed novel C2H2 zinc-finger protein with multiple
RT          consensus phosphorylation sites is conserved in mouse and man.";
RL          GENE 152:233-238(1995).
CC          -!- FUNCTION: CNTF IS A SURVIVAL FACTOR FOR VARIOUS NEURONAL CELL
CC          TYPES. SEEMS TO PREVENT THE DEGENERATION OF MOTOR AXONS AFTER
CC          AXOTOMY.
CC          -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC          -!- TISSUE SPECIFICITY: NERVOUS SYSTEM.
CC          -!- SIMILARITY: BELONGS TO THE CNTF FAMILY.
CC          -----
CC          This SWISS-PROT entry is copyright. It is produced through a collaboration
CC          between the Swiss Institute of Bioinformatics and the EMBL outstation
CC          the European Bioinformatics Institute. There are no restrictions on its
CC          use by non-profit institutions as long as its content is in no way
CC          modified and this statement is not removed. Usage by and for commercial
CC          entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC          or send an email to license@isb-sib.ch).
CC          -----
DR          EMBL; U05342; G453374; -.
DR          MGD; MGI:88439; CNTF.
DR          PFAM; PFO1110; CNTF; 1.
DR          HSSP; P26441; 1CNT.
DR          GROWTH FACTOR; NEURONE.
SQ          SEQUENCE      198 AA;  22587 MW;  A363A3A2 CRC32;

```

Query Match	1.58;	Score 8;	DB 1;	Length 198;
Best Local Similarity	100.0%;	Pred. No.	2.00e-02;	
Matches	8;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;

Db	50	LDSDVDGVP	57
QY	370	LDSDVDGVP	377

	RESULT	6	
ID	CNTF_RABIT	STANDARD;	PRT: 199 AA.
AC	P14188;		
DT	01-JAN-1990	(REL. 13, CREATED)	
DT	01-JAN-1990	(REL. 13, LAST SEQUENCE UPDATE)	
DT	01-OCT-1996	(REL. 34, LAST ANNOTATION UPDATE)	
DE	CILIARY NEUTROTROPHIC FACTOR (CNTF).		
GN	CNTF.		
OS	ORYCTOLAGUS CUNICULUS (RABBIT).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;		
OC	LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.		
RN	[1]		
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.		
RA	MEDLINE; 90069557.		
RA	LIN L.-F.H., MISMER D., LILE J.D., ARMES L.G., BUTLER E.T. III,		
RA	VANNICE J.L., COLLINS F.;		

OS CAENORHABDITIS VULGARIS.
OC EUKARYOTA; METAZOA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94061982.
RA YUAN J., SHAHAM S., LEDOUX S., ELLIS H.M., HORVITZ H.R.;
RT "The C. elegans cell death gene ced-3 encodes a protein similar to
RT mammalian interleukin-1 beta-converting enzyme.";
RL CELL 75:641-652(1993).
CC -1- FUNCTION: ACTS AS A CYSTEINE PROTEASE IN CONTROLLING PROGRAMMED
CC CELL DEATH BY PROTEOLYTICALLY ACTIVATING OR INACTIVATING A
CC SUBSTRATE PROTEIN OR PROTEINS. A POTENTIAL SUBSTRATE MAY BE CED-4.
CC ALTERNATIVELY IT MIGHT DIRECTLY CAUSE CELL DEATH BY
CC PROTEOLYTICALLY CLEAVING PROTEINS THAT ARE CRUCIAL FOR CELL
CC VIABILITY (BY SIMILARITY).
CC -1- SUBUNIT: COULD BE A HETERODIMER OF TWO SUBUNITS DERIVED FROM THE
CC PRECURSOR SEQUENCE BY A PROBABLE AUTOCATALYTIC MECHANISM.
CC -1- PTM: MAY BE REGULATED BY PHOSPHORYLATION.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CC CASPASE FAMILY.
UK PROSITE; PS01121; CASPASE_HIS; 1.
DR PROSITE; PS01122; CASPASE_CYS; 1.
DR PFAM; PF00619; CARD; 1.
DR PFAM; PF00655; ICE_P10; 1.
DR PFAM; PF00656; ICE_P20; 1.
DR HSSP; P42574; ICP3.
RV HYDROLASE; THIOLESTERASE; ZMOGEN; APOPTOSIS; PHOSPHORYLATION.
FT CHAIN 1 364
FT CHAIN 365 496
FT CHAIN 365 496
FT ACT_SITE 308 308
FT ACT_SITE 351 351
SO SEQUENCE 496 AA; 55945 MW; CBDFE84D CRC32;
Query Match 14.3%; Score 72; DB 1; Length 496;
Best Local Similarity 100.0%; Pred. No. 2.05e-222;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 394 QOYWRKKPSQADILIRYATTAQYVSWRNSARGSWFIQAVCEVSTHAKDMVYELLTEVN 453
QY 403 QOYWRKKPSQADILIRYATTAQYVSWRNSARGSWFIQAVCEVSTHAKDMVYELLTEVN 462
Db 454 KKVACGFQTSQG 465
QY 463 KKVACGFQTSQG 474
RESULT 3
ID ICE7_HUMAN STANDARD; PRT; 303 AA.
AC P55210; Q13364;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CASPASE-7 PRECURSOR (EC 3.4.22.-) (ICE-LIKE APOPTOTIC PROTEASE 3)
DE (ICE-LAP3) (APOPTOTIC PROTEASE MCH-3) (CMH-1).
GN CASP7 OR MCH3.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A. (ALPHA ISOFORM).
RX MEDLINE; 96139498.
RA DUAN H., CHINNATYAN A.M., HUDSON P.L., WING J.P., HE W.-W.,
RA DIXIT V.M.;
RT "ICE-LAP3, a novel mammalian homologue of the Caenorhabditis elegans
RT cell death protein Ced-3 is activated during Fas- and tumor necrosis
RT factor-induced apoptosis.";
RL J. BIOL. CHEM. 271:1621-1625(1996).
RN [2]
RP SEQUENCE FROM N.A. (ALPHA ISOFORM).
RC TISSUE=SPLEEN;

RX MEDLINE; 96147144.
RA LIPKKE J.A., GU Y., SARNECKI C., CARON P.R., SU M.S.-S.;
RT "Identification and characterization of CPP32/Mch2 homolog 1, a novel
RT cysteine protease similar to CPP32.";
RL J. BIOL. CHEM. 271:1825-1828(1996).
RN [3]
RP SEQUENCE FROM N.A. (ALPHA AND BETA ISOFORMS).
RC TISSUE=T-CELL;
RX MEDLINE; 96105019.
RA FERNANDES-ALNEMRI T., TAKAHASHI A., ARMSTRONG R.C., KREBS J.,
RA FRITZ L., TOMASELLI K.J., WANG L., YU Z., CROCE C.M., SALVESON G.,
RA FARNSHAW W.C., LITWACK G., ALNEMRI E.S.;
RT "Mch3, a novel human apoptotic cysteine protease highly related to
RT CPP32.";
RL CANCER RES. 55:6045-6052(1995).
RN [4]
RP SEQUENCE FROM N.A. (ALPHA AND ALPHA' ISOFORMS).
RC TISSUE=FETAL LUNG, AND FETAL SPLEEN;
RX MEDLINE; 97224489.
RA JUAN T.S.-C., MCNIECE I.K., ARGENTO J.M., JENKINS N.A., GILBERT D.J.,
RA COPELAND N.G., FLETCHER F.A.;
RT "Identification and mapping of Casp7, a cysteine protease resembling
RT CPP32 beta, interleukin-1 beta converting enzyme, and CED-3.";
RL GENOMICS 40:86-93(1997).
RN [5]
RP PROCESSING.
RX MEDLINE; 96353838.
RA FERNANDES-ALNEMRI T., ARMSTRONG R.C., KREBS J., SRINIVASULA S.M.,
RA WANG L., BULLRICH F., FRITZ L.C., TRAPANI J.A., TOMASELLI K.J.,
RA LITWACK G., ALNEMRI E.S.;
RT "In vitro activation of CPP32 and Mch3 by Mch4, a novel human
RT apoptotic cysteine protease containing two FADD-like domains.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:7464-7469(1996).
CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
CC RESPONSIBLE FOR APOPTOSIS EXECUTION. CLEAVES AND ACTIVATES STEROL
CC REGULATORY ELEMENT BINDING PROTEINS (SREBPS). PROTEOLYTICALLY
CC CLEAVES POLY(ADP-RIBOSE) POLYMERASE (PARP) AT A 216-ASP-1-GLY-217
CC BOND. OVEREXPRESSION PROMOTES PROGRAMMED CELL DEATH.
CC -1- SUBUNIT: HETERODIMER OF A 20 KD (P20) AND A 11 KD (P11) SUBUNIT.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG, SKELETAL MUSCLE,
CC LIVER, KIDNEY, SPLEEN AND HEART, AND MODERATELY IN TESTIS. NO
CC EXPRESSION IN THE BRAIN.
CC -1- PTM: CLEAVAGES BY GRANZYME B OR CASPASE-10 GENERATE THE TWO ACTIVE
CC SUBUNITS. PROPEPTIDE DOMAINS CAN ALSO BE CLEAVED EFFICIENTLY BY
CC CPP32 PROTEASE. ACTIVE HETERODIMERS BETWEEN THE SMALL SUBUNIT OF
CC CASPASE-7 AND THE LARGE SUBUNIT OF CPP32 PROTEASE ALSO OCCUR AND
CC VICE VERSA.
CC -1- ALTERNATIVE PRODUCTS: THREE ISOFORMS, ALPHA, BETA, AND ALPHA' ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THE SEQUENCE SHOWN HERE IS THAT
CC OF ALPHA ISOFORM. THE BETA ISOFORM IS NOT PROTEOLYTICALLY ACTIVE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
CC CASPASE FAMILY.
CC -1- CAUTION: WHAT WE CALL ALPHA' ISOFORM IS KNOWN IN REF.4 AS BETA,
CC BUT AS BETA IS ALREADY DEFINED IN REF.3 WE HAVE CALLED IT ALPHA'.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39613; G1125073; -.
DR EMBL; U40281; G1167555; -.
DR EMBL; U37448; G1117847; -.
DR EMBL; U37449; G1117849; -.
DR EMBL; U67319; G1894913; -.
DR EMBL; U67320; G1894915; -.
DR MIM; 601761; -.
DR PROSITE; PS01121; CASPASE_HIS; 1.
DR PROSITE; PS01122; CASPASE_CYS; 1.

973 5 1.0 1030 1 Y075_MYCPN HYPOTHETICAL PROTEIN M 3.29e+03
974 5 1.0 1076 1 YEM3_YEAST HYPOTHETICAL 119.3 KD 3.29e+03
975 5 1.0 1095 1 YOH5_YEAST PROBABLE ATP-DEPENDENT 3.29e+03
976 5 1.0 1105 1 YEGE_ECOLI HYPOTHETICAL 123.9 KD 3.29e+03
977 5 1.0 1139 1 VRNA_BSMV ALPHA-A PROTEIN. 3.29e+03
978 5 1.0 1159 1 Y196_HUMAN HYPOTHETICAL PROTEIN K 3.29e+03
979 5 1.0 1161 1 Y39P_YEAST HYPOTHETICAL 118.4 KD 3.29e+03
980 5 1.0 1165 1 YNF4_YEAST HYPOTHETICAL 128.1 KD 3.29e+03
981 5 1.0 1186 1 XPG_HUMAN DNA-REPAIR PROTEIN COM 3.29e+03
982 5 1.0 1192 1 XPG_MOUSE DNA-REPAIR PROTEIN COM 3.29e+03
983 5 1.0 1218 1 YK83_YEAST PROBABLE ATP-DEPENDENT 3.29e+03
984 5 1.0 1237 1 YD12_SCHPO HYPOTHETICAL 143.0 KD 3.29e+03
985 5 1.0 1251 1 YQ03_CAEL HYPOTHETICAL 133.5 KD 3.29e+03
986 5 1.0 1300 1 Y338_MYCPN HYPOTHETICAL LIPOPROTE 3.29e+03
987 5 1.0 1374 1 YAH8_SCHPO HYPOTHETICAL HELICASE 3.29e+03
988 5 1.0 1401 1 WRN_MOUSE WERNER SYNDROME HELICA 3.29e+03
989 5 1.0 1432 1 WRN_HUMAN WERNER SYNDROME HELICA 3.29e+03
990 5 1.0 1459 1 YF1M_CAEL HYPOTHETICAL 166.0 KD 3.29e+03
991 5 1.0 1478 1 YAWB_SCHPO PROBABLE ATP-DEPENDENT 3.29e+03
992 5 1.0 1517 1 YD22_SCHPO HYPOTHETICAL 170.7 KD 3.29e+03
993 5 1.0 1758 1 YIR7_YEAST HYPOTHETICAL 197.5 KD 3.29e+03
994 5 1.0 1764 1 YK84_YEAST HYPOTHETICAL 203.3 KD 3.29e+03
995 5 1.0 1888 1 YDT2_SCHPO HYPOTHETICAL 217.4 KD 3.29e+03
996 5 1.0 2054 1 YCF2_PINTH HYPOTHETICAL 244.6 KD 3.29e+03
997 5 1.0 2124 1 Y192_HUMAN HYPOTHETICAL PROTEIN K 3.29e+03
998 5 1.0 2280 1 YCF2_TOBAC HYPOTHETICAL 267 KD PR 3.29e+03
999 5 1.0 3051 1 YNX3_CAEL HYPOTHETICAL PROTEIN T 3.29e+03
1000 5 1.0 3655 1 YAMB_SCHPO HYPOTHETICAL 420.8 KD 3.29e+03

ALIGNMENTS

RESULT 1
ID CED3-CAEVL STANDARD; PRT; 503 AA.

AC P42573; P45435;

DT 01-NOV-1995 (REL. 32, CREATED)

DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE CELL DEATH PROTEIN 3 PRECURSOR (EC 3.4.22.-).

GN CED-3.

OS CAENORHABDITIS ELEGANS.

OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;

OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RX MEDLINE; 94061982.

RA YUAN J., SHAHAM S., LEDOUX S., ELLIS H.M., HORVITZ H.R.;

RT "the C. elegans cell death gene ced-3 encodes a protein similar to

RT mammalian interleukin-1 beta-converting enzyme.";

RL CELL 75:641-652(1993).

CC -1- FUNCTION: ACTS AS A CYSTEINE PROTEASE IN CONTROLLING PROGRAMMED

CC CELL DEATH BY PROTEOLYTICALLY ACTIVATING OR INACTIVATING A

CC SUBSTRATE PROTEIN OR PROTEINS, A POTENTIAL SUBSTRATE MAY BE CED-4.

CC ALTERNATIVELY IT MIGHT DIRECTLY CAUSE CELL DEATH BY

CC PROTEOLYTICALLY CLEAVING PROTEINS THAT ARE CRUCIAL FOR CELL

CC VIABILITY.

CC -1- SUBUNIT: COULD BE A HETERODIMER OF TWO SUBUNITS DERIVED FROM THE

CC PRECURSOR SEQUENCE BY A PROBABLE AUTOCATALYTIC MECHANISM.

CC -1- DEVELOPMENTAL STAGE: MOST ABUNDANT DURING EMBRYOGENESIS AND IS

CC ALSO DETECTED AT LATER STAGES.

CC -1- PTM: MAY BE REGULATED BY PHOSPHORYLATION.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE

CC CASPASE FAMILY.

CC -----

CC This Swiss-Prot entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

DR EMBL; L29052; G456417; -
DR PROSITE; PS01121; CASPASE_HIS; 1.
DR PROSITE; PS01122; CASPASE_CYS; 1.
DR PFAM; PF00619; CARD; 1.
DR PFAM; PF00655; ICE_p10; 1.
DR PFAM; PF00656; ICE_p20; 1.
DR HSSP; P42574; 1CP3.
KW HYDROLASE; THIOL PROTEASE; ZMOGEN; APOPTOSIS; PHOSPHORYLATION.
FT CHAIN 1 371
FT CHAIN 372 503
FT CHAIN 503
FT DOMAIN 107 205
FT ACT_SITE 304 304
FT ACT_SITE 358 358
FT MUTAGEN 27 27
FT MUTAGEN 65 65
FT MUTAGEN 360 360
FT MUTAGEN 449 449
FT MUTAGEN 466 466
FT MUTAGEN 483 483
FT MUTAGEN 486 486
SQ SEQUENCE 503 AA; 56702 MW; F6874EBE CRC32;

Query Match 100.0%; Score 503; DB 1; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MMRDRSLRLERNIMFSSHLKVDLEVLIAKQVLSNDGDMINSCTVREKREIYKA 60
QY 1 MMRDRSLRLERNIMFSSHLKVDLEVLIAKQVLSNDGDMINSCTVREKREIYKA 60
Db 61 VORRGDAFDALRSTGHEGLAEVLEPLARSVDSNAVEFECPMSPASHRRSRALSPA 120
QY 61 VORRGDAFDALRSTGHEGLAEVLEPLARSVDSNAVEFECPMSPASHRRSRALSPA 120
Db 121 GYTSPTRVHRDVSVSSTSYDYISRRASRSRSLHSSDRHNSPPVNAFPSPSS 180
QY 121 GYTSPTRVHRDVSVSSTSYDYISRRASRSRSLHSSDRHNSPPVNAFPSPSS 180
Db 181 ANSFTGSSISYSSSRNRSFSKASGPTQYTFHEEDMNFVDAPTISRVEDEKTMYNFSS 240
QY 181 ANSFTGSSISYSSSRNRSFSKASGPTQYTFHEEDMNFVDAPTISRVEDEKTMYNFSS 240
Db 241 PRGMCLINNEHEQMPTNRGTAKDKDNLNLFRCMGTYVICDKDNLTRGMLLTIRDFAK 300
QY 241 PRGMCLINNEHEQMPTNRGTAKDKDNLNLFRCMGTYVICDKDNLTRGMLLTIRDFAK 300
Db 301 HESHGDSAILVILSHGEENVIGVDIPISITHEIYDLLNAANAPRLANKPKIVFVQACRG 360
QY 301 HESHGDSAILVILSHGEENVIGVDIPISITHEIYDLLNAANAPRLANKPKIVFVQACRG 360
Db 361 EERDNGFPLDSDVGVPAFLRGWDNRDGLNLFGLCVRPQVQVWKKPSQADILIRYA 420
QY 361 EERDNGFPLDSDVGVPAFLRGWDNRDGLNLFGLCVRPQVQVWKKPSQADILIRYA 420
Db 421 TTAQVSWRNSARGSWFIQAVCEVFSTHAKMDVVELLTVNKKVACGFQTSQGSNLIKQ 480
QY 421 TTAQVSWRNSARGSWFIQAVCEVFSTHAKMDVVELLTVNKKVACGFQTSQGSNLIKQ 480
Db 481 MPEMTRLLKKFYFWPEARNSAV 503
QY 481 MPEMTRLLKKFYFWPEARNSAV 503
RESULT 2
ID CED3-CAEVU STANDARD; PRT; 496 AA.
AC P45436;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CELL DEATH PROTEIN 3 PRECURSOR (EC 3.4.22.-).

827	5	1.0	339	1	Y4FO_RHISN	PUTATIVE ABC TRANSPORT	3.29e+03	900	5	1.0	512	1	YES6_YEAST	HYPOTHETICAL 56.6 KD P	3.29e+03
828	5	1.0	341	1	VP3_GFLV	P3 PROTEIN.	3.29e+03	901	5	1.0	512	1	Y00K_MYCTU	HYPOTHETICAL 53.6 KD P	3.29e+03
829	5	1.0	342	1	YG1L_YEAST	HYPOTHETICAL 38.5 KD P	3.29e+03	902	5	1.0	519	1	VP6_MTVNJ	STRUCTURAL PROTEIN P6	3.29e+03
830	5	1.0	346	1	YQC2_CAEEL	HYPOTHETICAL 39.2 KD P	3.29e+03	903	5	1.0	534	1	YOG1_CAEEL	HYPOTHETICAL 59.2 KD P	3.29e+03
831	5	1.0	346	1	YNP1_CAEEL	HYPOTHETICAL 42.2 KD P	3.29e+03	904	5	1.0	542	1	ZYX_CHICK	ZYXIN.	3.29e+03
832	5	1.0	348	1	XKDT_BACSU	PHAGE-LIKE ELEMENT PBS	3.29e+03	905	5	1.0	543	1	ZN08_HUMAN	ZINC FINGER PROTEIN 8	3.29e+03
833	5	1.0	348	1	Z134_HUMAN	ZINC FINGER PROTEIN 13	3.29e+03	906	5	1.0	548	1	YDID_ECOLI	HYPOTHETICAL 60.8 KD P	3.29e+03
834	5	1.0	349	1	Y028_BORBU	HYPOTHETICAL PROTEIN B	3.29e+03	907	5	1.0	551	1	YIDE_HAEIN	HYPOTHETICAL PROTEIN H	3.29e+03
835	5	1.0	351	1	YIIG_ECOLI	HYPOTHETICAL 39.3 KD P	3.29e+03	908	5	1.0	559	1	YAEK_SCHPO	PUTATIVE INORGANIC PHO	3.29e+03
836	5	1.0	353	1	WN11_XENLA	WNT-11 PROTEIN PRECURS	3.29e+03	909	5	1.0	577	1	YGSU_SCHPO	HYPOTHETICAL 67.3 KD P	3.29e+03
837	5	1.0	359	1	Y4UE_RHISN	PUTATIVE TRANSPOSASE Y	3.29e+03	910	5	1.0	578	1	YAXB_SCHPO	PUTATIVE ATP-DEPENDENT	3.29e+03
838	5	1.0	360	1	Y303_HELPY	PROBABLE GTP-BINDING P	3.29e+03	911	5	1.0	584	1	YMB3_YEAST	HYPOTHETICAL 66.8 KD P	3.29e+03
839	5	1.0	360	1	YTPX_TTVL	VIRAL PROTEIN TPX.	3.29e+03	912	5	1.0	585	1	Y187_MYCGE	HYPOTHETICAL ABC TRANS	3.29e+03
840	5	1.0	360	1	Y0FA_CAEEL	HYPOTHETICAL 39.7 KD P	3.29e+03	913	5	1.0	597	1	YKP4_KULDA	HYPOTHETICAL KILLER PL	3.29e+03
841	5	1.0	361	1	Y101_NPVAC	HYPOTHETICAL 41.5 KD P	3.29e+03	914	5	1.0	599	1	YGM5_YEAST	HYPOTHETICAL 68.5 KD P	3.29e+03
842	5	1.0	361	1	YADE_SCHPO	HYPOTHETICAL 40.0 KD P	3.29e+03	915	5	1.0	599	1	YJ01_YEAST	HYPOTHETICAL 68.5 KD P	3.29e+03
843	5	1.0	366	1	Y433_ARCFU	HYPOTHETICAL PROTEIN A	3.29e+03	916	5	1.0	607	1	YSCC_YEREN	YOP PROTEINS TRANSLOCA	3.29e+03
844	5	1.0	368	1	Y4X3_CAEEL	HYPOTHETICAL 37.9 KD P	3.29e+03	917	5	1.0	614	1	Y3A3_SCHPO	HYPOTHETICAL 67.3 KD T	3.29e+03
845	5	1.0	369	1	YOK3_CAEEL	HYPOTHETICAL 41.7 KD P	3.29e+03	918	5	1.0	620	1	YEDX_THEMA	HYPOTHETICAL 72.9 KD P	3.29e+03
846	5	1.0	372	1	YJGP_HAEIN	HYPOTHETICAL PROTEIN H	3.29e+03	919	5	1.0	624	1	YE70_METJA	HYPOTHETICAL PROTEIN M	3.29e+03
847	5	1.0	373	1	YXX1_ACHAM	HYPOTHETICAL PROTEIN (3.29e+03	920	5	1.0	628	1	YKP5_YEAST	HYPOTHETICAL 72.2 KD P	3.29e+03
848	5	1.0	373	1	Y38K_THETE	HYPOTHETICAL 38 KD PRO	3.29e+03	921	5	1.0	631	1	YF14_HAEIN	HYPOTHETICAL PROTEIN H	3.29e+03
849	5	1.0	374	1	YN9C_YEAST	HYPOTHETICAL 43.9 KD P	3.29e+03	922	5	1.0	632	1	Y022_CAEEL	HYPOTHETICAL 73.8 KD P	3.29e+03
850	5	1.0	375	1	Y4ED_RHISN	HYPOTHETICAL 42.2 KD P	3.29e+03	923	5	1.0	642	1	YDIF_BACSU	HYPOTHETICAL ABC TRANS	3.29e+03
851	5	1.0	377	1	YTXN_BACSU	HYPOTHETICAL 42.9 KD P	3.29e+03	924	5	1.0	645	1	VP74_NPVCF	P74 PROTEIN.	3.29e+03
852	5	1.0	378	1	YK84_CAEEL	HYPOTHETICAL 43.0 KD P	3.29e+03	925	5	1.0	654	1	YPJ1_CAEEL	HYPOTHETICAL 75.1 KD P	3.29e+03
853	5	1.0	379	1	YFG2_YEAST	HYPOTHETICAL 45.3 KD P	3.29e+03	926	5	1.0	655	1	YJHG_ECOLI	HYPOTHETICAL 70.0 KD P	3.29e+03
854	5	1.0	381	1	YN06_CAEEL	HYPOTHETICAL 42.9 KD P	3.29e+03	927	5	1.0	660	1	Y390_MYCGE	HYPOTHETICAL ATP-BINDI	3.29e+03
855	5	1.0	385	1	YCT2_BACFI	HYPOTHETICAL 43.4 KD P	3.29e+03	928	5	1.0	660	1	Y390_MYCPN	HYPOTHETICAL ATP-BINDI	3.29e+03
856	5	1.0	390	1	YFGB_HAEIN	HYPOTHETICAL PROTEIN H	3.29e+03	929	5	1.0	661	1	Y05L_MYCTU	HYPOTHETICAL 70.4 KD P	3.29e+03
857	5	1.0	391	1	YACE_RHISN	HYPOTHETICAL 43.4 KD P	3.29e+03	930	5	1.0	661	1	WH13_YEAST	WH13 PROTEIN.	3.29e+03
858	5	1.0	395	1	YCHF_CAEEL	PUTATIVE GTP-BINDING P	3.29e+03	931	5	1.0	662	1	YFL8_YEAST	HYPOTHETICAL 75.9 KD P	3.29e+03
859	5	1.0	400	1	YNJ2_YEAST	HYPOTHETICAL 45.5 KD P	3.29e+03	932	5	1.0	663	1	Y104_METJA	HYPOTHETICAL ATP-BINDI	3.29e+03
860	5	1.0	401	1	YD76_METTM	HYPOTHETICAL 43.7 KD P	3.29e+03	933	5	1.0	670	1	YER8_YEAST	HYPOTHETICAL 73.0 KD P	3.29e+03
861	5	1.0	404	1	YJ9N_YEAST	HYPOTHETICAL 45.1 KD P	3.29e+03	934	5	1.0	676	1	Y043_YEAST	HYPOTHETICAL 77.0 KD P	3.29e+03
862	5	1.0	406	1	YDLB_SCHPO	HYPOTHETICAL 46.4 KD P	3.29e+03	935	5	1.0	683	1	YK86_CAEEL	HYPOTHETICAL 76.4 KD P	3.29e+03
863	5	1.0	406	1	Y503_METJA	PUTATIVE 2-ISOPROPYLMA	3.29e+03	936	5	1.0	683	1	YKB7_YEAST	HYPOTHETICAL 78.3 KD P	3.29e+03
864	5	1.0	407	1	YEAD_SCHPO	HYPOTHETICAL 45.4 KD G	3.29e+03	937	5	1.0	689	1	YBP3_YEAST	HYPOTHETICAL 77.3 KD P	3.29e+03
865	5	1.0	409	1	Y4CK_RHISN	PUTATIVE REPLICATION P	3.29e+03	938	5	1.0	702	1	Y4I4_YEAST	HYPOTHETICAL 78.3 KD P	3.29e+03
866	5	1.0	410	1	YD76_METTH	HYPOTHETICAL PROTEIN M	3.29e+03	939	5	1.0	715	1	Y07J_MYCTU	HYPOTHETICAL 78.2 KD P	3.29e+03
867	5	1.0	411	1	Y588_HAEIN	HYPOTHETICAL PROTEIN H	3.29e+03	940	5	1.0	722	1	Y290_HAEIN	PROBABLE CATION-TRANS	3.29e+03
868	5	1.0	414	1	YG35_METJA	HYPOTHETICAL PROTEIN M	3.29e+03	941	5	1.0	724	1	YCF4_YEAST	HYPOTHETICAL 83.2 KD P	3.29e+03
869	5	1.0	414	1	YCFW_ECOLI	HYPOTHETICAL 45.3 KD P	3.29e+03	942	5	1.0	724	1	YMF6_YEAST	HYPOTHETICAL 82.0 KD P	3.29e+03
870	5	1.0	417	1	YCHO_ECOLI	HYPOTHETICAL 47.0 KD P	3.29e+03	943	5	1.0	725	1	VR2A_BPT4	RIIA PROTEIN.	3.29e+03
871	5	1.0	424	1	Y146_MYCGE	HYPOTHETICAL PROTEIN M	3.29e+03	944	5	1.0	728	1	Y376_METJA	HYPOTHETICAL PROTEIN M	3.29e+03
872	5	1.0	431	1	Y4BN_RHISN	HYPOTHETICAL 47.6 KD P	3.29e+03	945	5	1.0	730	1	YEA5_SCHPO	HYPOTHETICAL 79.3 KD P	3.29e+03
873	5	1.0	431	1	YAHG_SCHPO	HYPOTHETICAL 48.4 KD P	3.29e+03	946	5	1.0	744	1	YGY4_HALSO	HYPOTHETICAL 80.2 KD P	3.29e+03
874	5	1.0	435	1	YBL6_YEAST	HYPOTHETICAL ALDEHYDE-	3.29e+03	947	5	1.0	751	1	YLX5_CAEEL	HYPOTHETICAL 84.8 KD P	3.29e+03
875	5	1.0	438	1	Y027_HUMAN	HYPOTHETICAL PROTEIN K	3.29e+03	948	5	1.0	759	1	YEHM_ECOLI	HYPOTHETICAL 83.4 KD P	3.29e+03
876	5	1.0	442	1	YJBI_ECOLI	HYPOTHETICAL 51.7 KD P	3.29e+03	949	5	1.0	770	1	YV05_MYCTU	PROBABLE CATION-TRANS	3.29e+03
877	5	1.0	447	1	Y4WB_RHISN	HYPOTHETICAL ZINC PROT	3.29e+03	950	5	1.0	772	1	YICL_ECOLI	HYPOTHETICAL 88.1 KD P	3.29e+03
878	5	1.0	448	1	Y329_MYCGE	HYPOTHETICAL 50.2 KD G	3.29e+03	951	5	1.0	782	1	XPB_HUMAN	DNA-REPAIR PROTEIN COM	3.29e+03
879	5	1.0	449	1	VP1_CAVCI	PROBABLE COAT PROTEIN	3.29e+03	952	5	1.0	783	1	XPB_MOUSE	DNA-REPAIR PROTEIN COM	3.29e+03
880	5	1.0	453	1	YEGQ_ECOLI	PUTATIVE PROTEASE IN B	3.29e+03	953	5	1.0	794	1	Y348_METJA	TAIL TUBULAR PROTEIN B	3.29e+03
881	5	1.0	457	1	ZXDA_HUMAN	PROTEIN U54.	3.29e+03	954	5	1.0	795	1	YTB_BPT7	DNA EXCISION REPAIR PR	3.29e+03
882	5	1.0	458	1	YC86_HAEIN	HYPOTHETICAL PROTEIN H	3.29e+03	955	5	1.0	802	1	YGN9_YEAST	HYPOTHETICAL 90.8 KD P	3.29e+03
883	5	1.0	459	1	Y030_NPVOP	HYPOTHETICAL 52.7 KD P	3.29e+03	956	5	1.0	802	1	YXDS_ECOLI	HYPOTHETICAL 92.8 KD P	3.29e+03
884	5	1.0	460	1	YEGQ_HAEIN	PUTATIVE PROTEASE H104	3.29e+03	957	5	1.0	807	1	YHX2_YEAST	HYPOTHETICAL 96.8 KD P	3.29e+03
885	5	1.0	461	1	Y061_MYCGE	HYPOTHETICAL PROTEIN M	3.29e+03	958	5	1.0	823	1	Y343_METJA	HYPOTHETICAL 105.3 KD	3.29e+03
886	5	1.0	462	1	YEBD_SCHPO	HYPOTHETICAL 53.5 KD P	3.29e+03	959	5	1.0	849	1	Y07C_MYCTU	HYPOTHETICAL ABC TRANS	3.29e+03
887	5	1.0	467	1	Y01J_CAEEL	PROBABLE AMINOTRANSFER	3.29e+03	960	5	1.0	859	1	YG72_HAEIN	HYPOTHETICAL PROTEIN H	3.29e+03
888	5	1.0	470	1	XNIF_XENLA	LOW MOLECULAR WEIGHT N	3.29e+03	961	5	1.0	881	1	YLB3_CAEEL	HYPOTHETICAL 100.9 KD	3.29e+03
889	5	1.0	472	1	YAHG_ECOLI	HYPOTHETICAL 50.4 KD P	3.29e+03	962	5	1.0	892	1	YMB6_YEAST	HYPOTHETICAL 99.5 KD P	3.29e+03
890	5	1.0	477	1	YGJ1_ECOLI	HYPOTHETICAL 52.1 KD P	3.29e+03	963	5	1.0	894	1	YB53_YEAST	HYPOTHETICAL 104.5 KD	3.29e+03
891	5	1.0	489	1	YK5_YEAST	HYPOTHETICAL 54.6 KD P	3.29e+03	964	5	1.0	924	1	YB53_YEAST	HYPOTHETICAL 103.2 KD	3.29e+03
892	5	1.0	490	1	YPLC_CLOPE	HYPOTHETICAL 55.7 KD P	3.29e+03	965	5	1.0	932	1	Y4I4_SCHPO	HYPOTHETICAL 103.2 KD	3.29e+03
893	5	1.0	491	1	YK5_YEAST	HYPOTHETICAL 56.1 KD P	3.29e+03	966	5	1.0	950	1	Y136_HUMAN	HYPOTHETICAL PROTEIN K	3.29e+03
894	5	1.0	491	1	YK5_YEAST	HYPOTHETICAL 56.1 KD P	3.29e+03	967	5	1.0	952	1	YK15_CAEEL	HYPOTHETICAL 105.3 KD	3.29e+03
895	5	1.0	497	1	YK5_YEAST	HYPOTHETICAL 56.1 KD P	3.29e+03	968	5	1.0	953	1	YK15_CAEEL	HYPOTHETICAL 109.8 KD	3.29e+03
896	5	1.0	497	1	YK5_YEAST	HYPOTHETICAL 56.1 KD P	3.29e+03	969	5	1.0	993	1	YD43_SCHPO	HYPOTHETICAL 111.9 KD	3.29e+03
897	5	1.0	499	1	YK5_YEAST	REGULATORY PROTEIN WET	3.29e+03	970	5	1.0	1002	1	YEMA_DROME	YEMANUCLEIN-ALPHA.	3.29e+03
898	5	1.0	499	1	YK5_YEAST	REGULATORY PROTEIN WET	3.29e+03	971	5	1.0	1005	1	YCF1_OENBE	HYPOTHETICAL PROTEIN (3.29e+03
899	5	1.0	510	1	YDID_SCHPO	HYPOTHETICAL 53.7 KD P	3.29e+03	972	5	1.0	1018	1	YK2_CAEEL	HYPOTHETICAL 115.2 KD	3.29e+03

681	6	1.2	1316	1	RPOC_MYCTU	DNA-DIRECTED RNA POLYM	1.29e+02	754	5	1.0	148	1	YGEE_ECOLI	HYPOTHETICAL 16.9 KD P	3.29e+03
682	6	1.2	1317	1	RPOD_SYNY3	DNA-DIRECTED RNA POLYM	1.29e+02	755	5	1.0	148	1	YCT9_YEAST	HYPOTHETICAL 16.0 KD P	3.29e+03
683	6	1.2	1319	1	MN1_HUMAN	PROBABLE TUMOR SUPPRES	1.29e+02	756	5	1.0	151	1	YDS8_SCHPO	HYPOTHETICAL 17.5 KD P	3.29e+03
684	6	1.2	1332	1	XKDO_BACSU	PHAGE-LIKE ELEMENT PBS	1.29e+02	757	5	1.0	154	1	YD88_SYNY3	HYPOTHETICAL 17.3 KD P	3.29e+03
685	6	1.2	1358	1	PUR4_YEAST	PHOSPHORIBOSYLFORMYLGL	1.29e+02	758	5	1.0	155	1	YFH3_YEAST	HYPOTHETICAL 18.2 KD P	3.29e+03
686	6	1.2	1386	1	Y064_MYCPN	HYPOTHETICAL PROTEIN M	1.29e+02	759	5	1.0	166	1	YELJ_SCHPO	HYPOTHETICAL 18.2 KD P	3.29e+03
687	6	1.2	1396	1	VLTF_BPT5	L-SHAPED TAIL FIBRE PR	1.29e+02	760	5	1.0	168	1	Y097_MYCTU	HYPOTHETICAL 18.2 KD P	3.29e+03
688	6	1.2	1402	1	SALM_DROVI	HOMOTIC PROTEIN SPALT	1.29e+02	761	5	1.0	175	1	Y433_METJA	HYPOTHETICAL 18.2 KD P	3.29e+03
689	6	1.2	1403	1	NKCR_HUMAN	NK-TUMOR RECOGNITION P	1.29e+02	762	5	1.0	178	1	Y4YJ_RHISN	HYPOTHETICAL 20.4 KD P	3.29e+03
690	6	1.2	1416	1	RPOC_TREPA	DNA-DIRECTED RNA POLYM	1.29e+02	763	5	1.0	189	1	YML6_PARTE	HYPOTHETICAL 22.4 KD P	3.29e+03
691	6	1.2	1518	1	KKK1_YEAST	PROBABLE SERINE/THREON	1.29e+02	764	5	1.0	191	1	Y645_METJA	HYPOTHETICAL 22.4 KD P	3.29e+03
692	6	1.2	1559	1	STCJ_EMENT	PUTATIVE STERIGMATOCYS	1.29e+02	765	5	1.0	192	1	YFM4_YEAST	HYPOTHETICAL 20.5 KD P	3.29e+03
693	6	1.2	1574	1	RPOC_AOUAE	DNA-DIRECTED RNA POLYM	1.29e+02	766	5	1.0	194	1	Y6A_SCHPO	HYPOTHETICAL 22.4 KD P	3.29e+03
694	6	1.2	1592	1	YHD5_YEAST	PROBABLE ATP-DEPENDENT	1.29e+02	767	5	1.0	195	1	YPEC_LACLC	HYPOTHETICAL 20.8 KD P	3.29e+03
695	6	1.2	1603	1	PSC_DROME	POSTERIOR SEX COMBS PR	1.29e+02	768	5	1.0	199	1	YW33_MYCTU	HYPOTHETICAL 20.8 KD P	3.29e+03
696	6	1.2	1659	1	VIT_ONCMY	VITELLOGENIN PRECURSOR	1.29e+02	769	5	1.0	203	1	YBM4_YEAST	HYPOTHETICAL 22.6 KD P	3.29e+03
697	6	1.2	1676	1	APSA_EMENT	ANUCLEATE PRIMARY STER	1.29e+02	770	5	1.0	203	1	Y296_METJA	HYPOTHETICAL 22.6 KD P	3.29e+03
698	6	1.2	1683	1	YJD9_YEAST	HYPOTHETICAL 191.5 KD	1.29e+02	771	5	1.0	209	1	YCR7_YEAST	HYPOTHETICAL RAS-RELAT	3.29e+03
699	6	1.2	1698	1	Y076_HUMAN	HYPOTHETICAL PROTEIN K	1.29e+02	772	5	1.0	209	1	Y139_YEAST	HYPOTHETICAL PROTEIN M	3.29e+03
700	6	1.2	1722	1	RBB2_HUMAN	RETINOBLASTOMA BINDING	1.29e+02	773	5	1.0	210	1	YIHA_ECOLI	HYPOTHETICAL GTP-BINDI	3.29e+03
701	6	1.2	1729	1	RRP5_YEAST	RRNA BIOGENESIS PROTEI	1.29e+02	774	5	1.0	212	1	YMF3_YEAST	HYPOTHETICAL 23.9 KD P	3.29e+03
702	6	1.2	1736	1	Z01_HUMAN	TIGHT JUNCTION PROTEIN F	1.29e+02	775	5	1.0	213	1	YMO6_YEAST	HYPOTHETICAL 24.6 KD P	3.29e+03
703	6	1.2	1745	1	YL51_CAEEL	HYPOTHETICAL PROTEIN F	1.29e+02	776	5	1.0	218	1	YTXE_BACME	HYPOTHETICAL 24.6 KD P	3.29e+03
704	6	1.2	1786	1	LMB1_HUMAN	LAMININ BETA-1 CHAIN P	1.29e+02	777	5	1.0	218	1	Y556_SYNY3	HYPOTHETICAL 24.0 KD P	3.29e+03
705	6	1.2	1805	1	HMW2_MYCGE	CYTADHERENCE HIGH MOLE	1.29e+02	778	5	1.0	219	1	YGIX_ECOLI	PROBABLE TRANSCRIPTION	3.29e+03
706	6	1.2	1818	1	HMW2_MYCPN	CYTADHERENCE HIGH MOLE	1.29e+02	779	5	1.0	225	1	XYN1_EMENT	ENDO-1,4-BETA-XYLANASE	3.29e+03
707	6	1.2	2194	1	SC16_YEAST	MULTIDOMAIN VESICLE CO	1.29e+02	780	5	1.0	226	1	Y247_MYCGE	PROBABLE AMINO-ACID AB	3.29e+03
708	6	1.2	2209	1	Y166_HUMAN	HYPOTHETICAL PROTEIN K	1.29e+02	781	5	1.0	239	1	YBGP_ECOLI	HYPOTHETICAL FIMBRIAL	3.29e+03
709	6	1.2	2252	1	POL1_TBRVS	RNA1 POLYPROTEIN (250	1.29e+02	782	5	1.0	242	1	YBPA_BURCE	HYPOTHETICAL TRANSCRIP	3.29e+03
710	6	1.2	2264	1	POL1_GCMV	RNA1 POLYPROTEIN (250	1.29e+02	783	5	1.0	245	1	Y07E_MYCTU	PUTATIVE OXIDOREDUCTAS	3.29e+03
711	6	1.2	2325	1	FRAP_RAT	CHONDROITIN SULFATE PR	1.29e+02	784	5	1.0	247	1	Y085_YEAST	HYPOTHETICAL 27.6 KD P	3.29e+03
712	6	1.2	2549	1	FRAP_HUMAN	EKBP-RAPAMYCIN ASSOCIA	1.29e+02	785	5	1.0	249	1	YF57_METJA	HYPOTHETICAL 27.6 KD P	3.29e+03
713	6	1.2	2594	1	7LES_DROVI	SEVENLESS PROTEIN (EC	1.29e+02	786	5	1.0	251	1	Y4EF_RHISN	PUTATIVE INTEGRASE/REC	3.29e+03
714	6	1.2	2763	1	TEGU_VZVD	LARGE TEGUMENT PROTEIN	1.29e+02	787	5	1.0	255	1	YHJH_ECOLI	HYPOTHETICAL 29.6 KD P	3.29e+03
715	6	1.2	3075	1	LMAL_HUMAN	LAMININ ALPHA-1 CHAIN	1.29e+02	788	5	1.0	261	1	Y046_MYCTU	HYPOTHETICAL TRANSCRIP	3.29e+03
716	6	1.2	3084	1	HD_HUMAN	HUNTINGTIN (HUNTINGTON	1.29e+02	789	5	1.0	261	1	Y046_MYCTU	HYPOTHETICAL PROTEIN M	3.29e+03
717	6	1.2	3144	1	RPOA_EAV	POL POLYPROTEIN (ORF1A	1.29e+02	790	5	1.0	262	1	YGE6_YEAST	HYPOTHETICAL 29.4 KD P	3.29e+03
718	6	1.2	3175	1	POLG_DEN4	GENOME POLYPROTEIN (CO	1.29e+02	791	5	1.0	262	1	YHFJ_SALTY	HYPOTHETICAL 26.3 KD P	3.29e+03
719	6	1.2	3386	1	OL56_STRAT	OLENOMYCIN POLYKETID	1.29e+02	792	5	1.0	264	1	Y441_METJA	HYPOTHETICAL PROTEIN M	3.29e+03
720	6	1.2	3519	1	ANKB_HUMAN	ANKRYRIN, BRAIN VARIANT	1.29e+02	793	5	1.0	267	1	YDLD_ECOLI	HYPOTHETICAL OXIDORED	3.29e+03
721	6	1.2	3924	1	PGBM_HUMAN	BASEMENT MEMBRANE-SPEC	1.29e+02	794	5	1.0	269	1	YDLO_SCHPO	HYPOTHETICAL 29.9 KD P	3.29e+03
722	6	1.2	4393	1	PKSL_BACSU	PUTATIVE POLYKETIDE SY	1.29e+02	795	5	1.0	270	1	YDIJ_SCHPO	HYPOTHETICAL 31.5 KD P	3.29e+03
723	6	1.2	4427	1	APB_HUMAN	APOLIPROTEIN B-100 P	1.29e+02	796	5	1.0	275	1	YD84_YEAST	HYPOTHETICAL 30.0 KD P	3.29e+03
724	6	1.2	4563	1	YCL8_AGLNE	HYPOTHETICAL 7.0 KD PR	3.29e+03	797	5	1.0	275	1	Y284_CHVP1	HYPOTHETICAL PROTEIN A	3.29e+03
725	5	1.0	57	1	YCF9_CYAPA	HYPOTHETICAL 6.9 KD PR	3.29e+03	798	5	1.0	282	1	Y32K_BNYVG	RNA-4 HYPOTHETICAL 31.	3.29e+03
726	5	1.0	65	1	YEDF_ECOLI	HYPOTHETICAL 8.6 KD PR	3.29e+03	799	5	1.0	286	1	YMD3_CAEEL	HYPOTHETICAL 32.1 KD P	3.29e+03
727	5	1.0	77	1	YE96_HAEIN	HYPOTHETICAL PROTEIN H	3.29e+03	800	5	1.0	286	1	Y137_METJA	HYPOTHETICAL PROTEIN M	3.29e+03
728	5	1.0	84	1	YGIU_ECOLI	HYPOTHETICAL 11.2 KD P	3.29e+03	801	5	1.0	286	1	YCX1_ASTLO	HYPOTHETICAL 34.5 KD P	3.29e+03
729	5	1.0	98	1	Y4CB_RHISN	HYPOTHETICAL 10.2 KD P	3.29e+03	802	5	1.0	287	1	Y32K_SSV1	HYPOTHETICAL 31.7 KD P	3.29e+03
730	5	1.0	98	1	Y073_NPVAC	HYPOTHETICAL 11.5 KD P	3.29e+03	803	5	1.0	291	1	YG47_HAEIN	HYPOTHETICAL PROTEIN H	3.29e+03
731	5	1.0	99	1	YK85_CAEEL	HYPOTHETICAL 11.7 KD P	3.29e+03	804	5	1.0	291	1	YCUO_ECOLI	HYPOTHETICAL ABC TRANS	3.29e+03
732	5	1.0	102	1	Y18K_MSVK	HYPOTHETICAL 12.5 KD P	3.29e+03	805	5	1.0	293	1	YFEC_YERPE	HYPOTHETICAL TRANSRIP	3.29e+03
733	5	1.0	108	1	YCXD_PORPU	HYPOTHETICAL 12.2 KD P	3.29e+03	806	5	1.0	294	1	YFEC_YERPE	CHELATED IRON TRANSPO	3.29e+03
734	5	1.0	110	1	Y122_CAEEL	HYPOTHETICAL 12.2 KD P	3.29e+03	807	5	1.0	294	1	YFEC_YERPE	HYPOTHETICAL OXIDORED	3.29e+03
735	5	1.0	111	1	Y126_ADE07	HYPOTHETICAL 12.6 KD E	3.29e+03	808	5	1.0	294	1	YFEC_YERPE	HYPOTHETICAL 34.8 KD P	3.29e+03
736	5	1.0	112	1	Y126_ADE07	HYPOTHETICAL 12.6 KD E	3.29e+03	809	5	1.0	299	1	YFEC_YERPE	MATRIX PROTEIN VP40 (V	3.29e+03
737	5	1.0	114	1	Y126_ADE07	HYPOTHETICAL 12.6 KD E	3.29e+03	810	5	1.0	303	1	YDS4_SCHPO	HYPOTHETICAL 35.8 KD P	3.29e+03
738	5	1.0	121	1	Y102_NPVAC	HYPOTHETICAL 13.3 KD P	3.29e+03	811	5	1.0	306	1	YAHB_ECOLI	HYPOTHETICAL TRANSCRIP	3.29e+03
739	5	1.0	122	1	YD1C_ECOLI	HYPOTHETICAL 13.3 KD P	3.29e+03	812	5	1.0	310	1	YFM4_MICOL	HYPOTHETICAL PROTEIN I	3.29e+03
740	5	1.0	122	1	WNT7_EVATR	WNT-7 PROTEIN (FRAGMEN	3.29e+03	813	5	1.0	313	1	YX11_CAEEL	HYPOTHETICAL 35.5 KD P	3.29e+03
741	5	1.0	123	1	WNT7_EVATR	WNT-7 PROTEIN (FRAGMEN	3.29e+03	814	5	1.0	315	1	YX11_CAEEL	HYPOTHETICAL 35.5 KD P	3.29e+03
742	5	1.0	123	1	WNT7_EVATR	WNT-7 PROTEIN (FRAGMEN	3.29e+03	815	5	1.0	315	1	YX11_CAEEL	HYPOTHETICAL 35.5 KD P	3.29e+03
743	5	1.0	123	1	WNT7_EVATR	WNT-7 PROTEIN (FRAGMEN	3.29e+03	816	5	1.0	315	1	YX11_CAEEL	HYPOTHETICAL 35.5 KD P	3.29e+03
744	5	1.0	125	1	Y02D_BPT4	HYPOTHETICAL 13.9 KD P	3.29e+03	817	5	1.0	316	1	YX11_CAEEL	HYPOTHETICAL 35.5 KD P	3.29e+03
745	5	1.0	131	1	YJG4_YEAST	HYPOTHETICAL 13.9 KD P	3.29e+03	818	5	1.0	317	1	YX11_CAEEL	HYPOTHETICAL 35.5 KD P	3.29e+03
746	5	1.0	137	1	Y0HC_MYCTU	HYPOTHETICAL 15.0 KD P	3.29e+03	819	5	1.0	318	1	YX11_CAEEL	HYPOTHETICAL 35.5 KD P	3.29e+03
747	5	1.0	140	1	YJ14_YEAST	HYPOTHETICAL 15.7 KD P	3.29e+03	820	5	1.0	324	1	YB61_YEAST	HYPOTHETICAL 37.5 KD P	3.29e+03
748	5	1.0	142	1	YKAI_SCHPO	HYPOTHETICAL PROTEIN C	3.29e+03	821	5	1.0	325	1	YB61_YEAST	HYPOTHETICAL 37.5 KD P	3.29e+03
749	5	1.0	144	1	Y850_HAEIN	HYPOTHETICAL PROTEIN H	3.29e+03	822	5	1.0	325	1	YB61_YEAST	HYPOTHETICAL 37.5 KD P	3.29e+03
750	5	1.0	146	1	YAN3_SCHPO	HYPOTHETICAL 17.2 KD P	3.29e+03	823	5	1.0	330	1	Y4KL_RHISN	VP6 PROTEIN (MINOR INN	3.29e+03
751	5	1.0	147	1	YJV1_YEAST	VERY HYPOTHETICAL 16.8	3.29e+03	824	5	1.0	332	1	YQO1_CAEEL	HYPOTHETICAL 37.6 KD A	3.29e+03
752	5	1.0	148	1	YEV0_YEAST	HYPOTHETICAL 15.0 KD P	3.29e+03	825	5	1.0	336	1	YNV3_CAEEL	HYPOTHETICAL 37.5 KD P	3.29e+03
753	5	1.0	148	1	YFB2_YEAST	HYPOTHETICAL 17.2 KD P	3.29e+03	826	5	1.0	338	1	YABT_BACSU	PROBABLE SERINE/THREON	3.29e+03

535	6	1.2	595	1	YPD7_CAEEL	HYPOTHETICAL 66.7 KD P	1.29e+02	608	6	1.2	840	1	CC16_YEAST	CELL DIVISION CONTROL	1.29e+02
536	6	1.2	599	1	TPA_HELPY	GTP-BINDING PROTEIN TY	1.29e+02	609	6	1.2	843	1	CYPL_BRUMA	PEPTIDYLPROLYL ISOMERA	1.29e+02
537	6	1.2	601	1	OAR_DROME	OCTOPAMINE RECEPTOR PR	1.29e+02	610	6	1.2	853	1	PBPA_HAEIN	PENICILLIN-BINDING PRO	1.29e+02
538	6	1.2	603	1	PLK1_HUMAN	SERINE/THREONINE-PROTE	1.29e+02	611	6	1.2	859	1	ENV_EIAYV	ENV POLYPROTEIN PRECUR	1.29e+02
539	6	1.2	603	1	PRIM_BACSU	DNA PRIMASE (EC 2.7.7.	1.29e+02	612	6	1.2	859	1	ENV_EIAYV	ENV POLYPROTEIN PRECUR	1.29e+02
540	6	1.2	603	1	PLK1_MOUSE	SERINE/THREONINE-PROTE	1.29e+02	613	6	1.2	859	1	ENV_EIAYV	ENV POLYPROTEIN PRECUR	1.29e+02
541	6	1.2	603	1	LEPA_SYNY3	GTP-BINDING PROTEIN LE	1.29e+02	614	6	1.2	859	1	ENV_EIAYV	ENV POLYPROTEIN PRECUR	1.29e+02
542	6	1.2	605	1	AOR_PYREF	TUNGSTEN-CONTAINING AL	1.29e+02	615	6	1.2	859	1	ENV_EIAYV	ENV POLYPROTEIN PRECUR	1.29e+02
543	6	1.2	607	1	DB10_NICSY	RNA HELICASE-LIKE PROT	1.29e+02	616	6	1.2	859	1	ENV_EIAYV	ENV POLYPROTEIN PRECUR	1.29e+02
544	6	1.2	609	1	YAS8_METJA	HYPOTHETICAL PROTEIN M	1.29e+02	617	6	1.2	860	1	ENV_EIAYV	ENV POLYPROTEIN PRECUR	1.29e+02
545	6	1.2	614	1	RUI7_HUMAN	UT SMALL NUCLEAR RIBON	1.29e+02	618	6	1.2	860	1	ENV_EIAYV	ENV POLYPROTEIN PRECUR	1.29e+02
546	6	1.2	617	1	HEMA_MEASI	HEMAGGLUTININ-NEURAMIN	1.29e+02	619	6	1.2	868	1	CLP_TRYBB	HEAT SHOCK PROTEIN 100	1.29e+02
547	6	1.2	617	1	PYS1_PSEAE	PYOCIN SI (EC 3.1.21.1	1.29e+02	620	6	1.2	868	1	MISS_SCHPO	MISS PROTEIN.	1.29e+02
548	6	1.2	619	1	OM70_NEUCR	MITOCHONDRIAL PRECURSO	1.29e+02	621	6	1.2	869	1	NIB80_YEAST	NIB80 PROTEIN (NIP100	1.29e+02
549	6	1.2	622	1	YRT1_CAEEL	HYPOTHETICAL 71.6 KD P	1.29e+02	622	6	1.2	869	1	RIR3_YEAST	RIBONUCLEOSIDE-DIPHOSP	1.29e+02
550	6	1.2	622	1	YAG1_SCHPO	HYPOTHETICAL PROTEIN C	1.29e+02	623	6	1.2	875	1	LON1_SPIOI	MITOCHONDRIAL LON PROT	1.29e+02
551	6	1.2	630	1	TNPB_STAUI	TRANSPOSASE B (TRANSP	1.29e+02	624	6	1.2	879	1	MVSP_DROME	PARAMYOSIN, LONG FORM.	1.29e+02
552	6	1.2	630	1	TNPE_STAUI	TRANSPOSASE B (TRANSP	1.29e+02	625	6	1.2	885	1	LON1_MAIZE	MITOCHONDRIAL LON PROT	1.29e+02
553	6	1.2	632	1	YK65_CAEEL	HYPOTHETICAL 71.9 KD P	1.29e+02	626	6	1.2	891	1	POL2_BAMMN	GENOME POLYPROTEIN 2 (1.29e+02
554	6	1.2	637	1	YDOD_SCHPO	HYPOTHETICAL 71.6 KD P	1.29e+02	627	6	1.2	892	1	TRAI_STRFR	TRANSPOSASE FOR TRANSP	1.29e+02
555	6	1.2	638	1	PAN1_RAT	TRANSCRIPTIONAL REGULA	1.29e+02	628	6	1.2	905	1	GYRA_RICPR	DNA GYRASE SUBUNIT A (1.29e+02
556	6	1.2	639	1	PRIM_MYCTU	DNA PRIMASE (EC 2.7.7.	1.29e+02	629	6	1.2	907	1	NUOG_ECOLI	NADH DEHYDROGENASE I C	1.29e+02
557	6	1.2	646	1	UVRB_METTH	EXCINUCLEASE ABC SUBUN	1.29e+02	630	6	1.2	907	1	NUOG_SALTY	NADH DEHYDROGENASE I C	1.29e+02
558	6	1.2	647	1	BGAL_MOUSE	BETA-GALACTOSIDASE PRE	1.29e+02	631	6	1.2	917	1	VGLB_HSVB2	GLYCOPROTEIN B-1 PRECU	1.29e+02
559	6	1.2	649	1	PAN2_RAT	TRANSCRIPTIONAL REGULA	1.29e+02	632	6	1.2	921	1	RB_MOUSE	RETINOBLASTOMA-ASSOCIA	1.29e+02
560	6	1.2	652	1	FAS1_DROME	FASCICLIN I PRECURSOR	1.29e+02	633	6	1.2	928	1	ATC1_YARLI	CALCIUM-TRANSPORTING A	1.29e+02
561	6	1.2	663	1	GR78_SCHPO	78 KD GLUCOSE REGULATE	1.29e+02	634	6	1.2	928	1	RB_HUMAN	RETINOBLASTOMA-ASSOCIA	1.29e+02
562	6	1.2	668	1	YBFL_YEAST	HYPOTHETICAL 73.8 KD P	1.29e+02	635	6	1.2	935	1	YJ31_YEAST	HYPOTHETICAL 108.4 KD	1.29e+02
563	6	1.2	669	1	FPS1_YEAST	GLYCEROL UPTAKE/EFFLUX	1.29e+02	636	6	1.2	962	1	GUNA_PSEFL	ENDOGUCANASE A PRECUR	1.29e+02
564	6	1.2	673	1	YNM5_YEAST	HYPOTHETICAL 73.8 KD P	1.29e+02	637	6	1.2	979	1	PTPN_MOUSE	PROTEIN-TYROSINE PHOSP	1.29e+02
565	6	1.2	676	1	KR74_HSV11	GENE 74 PROTEIN KINASE	1.29e+02	638	6	1.2	979	1	PTPN_HUMAN	PROTEIN-TYROSINE PHOSP	1.29e+02
566	6	1.2	676	1	UL06_HSV11	VIRION PROTEIN UL6.	1.29e+02	639	6	1.2	983	1	PTPN_RAT	PROTEIN-TYROSINE PHOSP	1.29e+02
567	6	1.2	677	1	FLHA_BACSU	FLAGELLAR BIOSYNTHESIS	1.29e+02	640	6	1.2	1020	1	NFH_HUMAN	NEUROFILAMENT TRIPLET	1.29e+02
568	6	1.2	681	1	TRE_MANSE	TRANSFERRIN PRECURSOR.	1.29e+02	641	6	1.2	1021	1	MANA_RHOMR	MANNAN ENDO-1,4-BETA-M	1.29e+02
569	6	1.2	682	1	PLI3_PSEAE	PIL3 PROTEIN.	1.29e+02	642	6	1.2	1030	1	PEX6_YEAST	PEROXISOME BIOSYNTHESI	1.29e+02
570	6	1.2	685	1	SNWA_DICDI	SNWA PROTEIN.	1.29e+02	643	6	1.2	1042	1	SUMA_DROME	SUPPRESSOR OF WHITE AP	1.29e+02
571	6	1.2	690	1	ATKB_SYNY3	POTASSIUM-TRANSPORTING	1.29e+02	644	6	1.2	1046	1	YJ14_YEAST	HYPOTHETICAL 117.2 KD	1.29e+02
572	6	1.2	691	1	HSE_DROME	HEAT SHOCK FACTOR PROT	1.29e+02	645	6	1.2	1065	1	CERU_HUMAN	CERULOPLASMIN PRECURSO	1.29e+02
573	6	1.2	695	1	XET_HUMAN	PROTEIN XET.	1.29e+02	646	6	1.2	1073	1	Y126_HUMAN	HYPOTHETICAL PROTEIN K	1.29e+02
574	6	1.2	700	1	PALY_WHEAT	PHENYLALANINE AMONIA-	1.29e+02	647	6	1.2	1078	1	C27A_MOUSE	MHC CLASS II TRANSACT B	1.29e+02
575	6	1.2	702	1	CCEM_HUMAN	CARCINOEMBRYONIC ANTIG	1.29e+02	648	6	1.2	1078	1	GYRB_SYNY3	DNA GYRASE SUBUNIT B (1.29e+02
576	6	1.2	702	1	COAT_SMSV1	COAT PROTEIN (CAPSID P	1.29e+02	649	6	1.2	1081	1	UL52_HSVB	DNA HELICASE/PRIMASE C	1.29e+02
577	6	1.2	705	1	YMTF_BACSU	HYPOTHETICAL 79.2 KD P	1.29e+02	650	6	1.2	1086	1	NNTM_BOVIN	NAD(P) TRANSHYDROGENAS	1.29e+02
578	6	1.2	707	1	DCOR_LEIDO	ORNITHINE DECARBOXYLAS	1.29e+02	651	6	1.2	1087	1	NFH_MOUSE	NEUROFILAMENT TRIPLET	1.29e+02
579	6	1.2	708	1	Y253_HUMAN	HYPOTHETICAL PROTEIN K	1.29e+02	652	6	1.2	1091	1	DIA_DROME	DIAPHANOUS PROTEIN.	1.29e+02
580	6	1.2	722	1	MAS2_ECOLI	MALATE SYNTHASE G (EC	1.29e+02	653	6	1.2	1092	1	DPOD_DROME	DNA POLYMERASE DELTA C	1.29e+02
581	6	1.2	727	1	IF2M_BOVIN	TRANSLATION INITIATION	1.29e+02	654	6	1.2	1093	1	AF17_HUMAN	AF-17 PROTEIN.	1.29e+02
582	6	1.2	731	1	TOP1_TREPA	DNA TOPOISOMERASE I (E	1.29e+02	655	6	1.2	1097	1	IMB3_HUMAN	IMPORTIN BETA-3 SUBUNI	1.29e+02
583	6	1.2	738	1	MASY_CORGL	MALATE SYNTHASE (EC 4.	1.29e+02	656	6	1.2	1111	1	YUEP_HAEIN	HYPOTHETICAL PROTEIN H	1.29e+02
584	6	1.2	739	1	UL47_HSVB	80.7 KD ALPHA TRANS-IN	1.29e+02	657	6	1.2	1128	1	PHY4_AVEA	PHYTOCHROME A TYPE 3 (1.29e+02
585	6	1.2	741	1	MASY_MYCTU	PROBABLE MALATE SYNTHA	1.29e+02	658	6	1.2	1128	1	PHY4_AVEA	PHYTOCHROME A TYPE 4 (1.29e+02
586	6	1.2	742	1	ZFA_MOUSE	ZINC FINGER AUTOSOMAL	1.29e+02	659	6	1.2	1130	1	C27A_HUMAN	MHC CLASS II TRANSACTI	1.29e+02
587	6	1.2	742	1	UL47_HSVBP	80.7 KD ALPHA TRANS-IN	1.29e+02	660	6	1.2	1132	1	BAT3_HUMAN	LARGE PROLINE-RICH PRO	1.29e+02
588	6	1.2	743	1	YK47_YEAST	HYPOTHETICAL 83.6 KD P	1.29e+02	661	6	1.2	1132	1	NUT1_YEAST	NUT1 PROTEIN.	1.29e+02
589	6	1.2	747	1	KF3B_MOUSE	KINESIN-LIKE PROTEIN K	1.29e+02	662	6	1.2	1139	1	SRE2_CRIGR	STEROL REGULATORY ELEM	1.29e+02
590	6	1.2	747	1	KF3B_HUMAN	KINESIN-LIKE PROTEIN K	1.29e+02	663	6	1.2	1141	1	SRE2_HUMAN	STEROL REGULATORY ELEM	1.29e+02
591	6	1.2	747	1	CLC4_MOUSE	CHLORIDE CHANNEL PROTE	1.29e+02	664	6	1.2	1142	1	PAK1_YEAST	SERINE/THREONINE-PROTE	1.29e+02
592	6	1.2	749	1	SWAP_CAEEL	SWAP PROTEIN (SUPPRESS	1.29e+02	665	6	1.2	1184	1	POL2_GFLV	RNA2 POLYPROTEIN (131	1.29e+02
593	6	1.2	752	1	EEGM_RAT	ELONGATION FACTOR G, M	1.29e+02	666	6	1.2	1199	1	RPOC_BACSU	DNA-DIRECTED RNA POLYM	1.29e+02
594	6	1.2	758	1	CHLD_TOBAC	MAGNESIUM-CHELATASE SU	1.29e+02	667	6	1.2	1209	1	DNBI_HSVB	MAJOR DNA-BINDING PROT	1.29e+02
595	6	1.2	760	1	DP4_MOUSE	DIPEPTIDYL PEPTIDASE I	1.29e+02	668	6	1.2	1222	1	YMH5_CAEEL	HYPOTHETICAL 136.3 KD	1.29e+02
596	6	1.2	770	1	LEU2_CANMA	3-ISOPROPYLMALATE DEHY	1.29e+02	669	6	1.2	1224	1	RPOD_PORPU	DNA-DIRECTED RNA POLYM	1.29e+02
597	6	1.2	773	1	PAC2_PSE3	PENICILLIN ACYLASE II	1.29e+02	670	6	1.2	1228	1	SLAP_BACST	S-LAYER PROTEIN PRECUR	1.29e+02
598	6	1.2	773	1	FE0B_ECOLI	FERRIC IRON TRANSPORT	1.29e+02	671	6	1.2	1254	1	DPOC_YEAST	DNA POLYMERASE GAMMA (1.29e+02
599	6	1.2	776	1	VP4_ROTSS	OUTER CAPSID PROTEIN V	1.29e+02	672	6	1.2	1265	1	RPOD_CYAPA	DNA-DIRECTED RNA POLYM	1.29e+02
600	6	1.2	776	1	VP42_ROTSS	OUTER CAPSID PROTEIN V	1.29e+02	673	6	1.2	1272	1	UBP2_YEAST	UBIQUITIN CARBOXYL-TER	1.29e+02
601	6	1.2	781	1	YB68_SCHPO	DYNAMIN-LIKE PROTEIN C	1.29e+02	674	6	1.2	1286	1	SK12_YEAST	ANTIVIRAL PROTEIN SK12	1.29e+02
602	6	1.2	781	1	KRAF_DROME	RAF HOMOLOG SERINE/THR	1.29e+02	675	6	1.2	1286	1	RPOD_GUITH	DNA-DIRECTED RNA POLYM	1.29e+02
603	6	1.2	798	1	YLFN_CAEEL	HYPOTHETICAL 89.8 KD P	1.29e+02	676	6	1.2	1289	1	VG34_BPT4	LARGE TAIL FIBER PROTE	1.29e+02
604	6	1.2	812	1	PLMN_BOVIN	PLASMINOGEN PRECURSOR	1.29e+02	677	6	1.2	1290	1	RPOC_MYCPN	DNA-DIRECTED RNA POLYM	1.29e+02
605	6	1.2	815	1	ERK5_HUMAN	EXTRACELLULAR SIGNAL-R	1.29e+02	678	6	1.2	1292	1	RPOC_MYCGE	DNA-DIRECTED RNA POLYM	1.29e+02
606	6	1.2	824	1	PML3_HUMAN	PROBABLE TRANSCRIPTION	1.29e+02	679	6	1.2	1299	1	VP75_HSVSA	PROBABLE MEMBRANE ANTI	1.29e+02
607	6	1.2	832	1	KLP1_SCHPO	KINESIN-LIKE PROTEIN 1	1.29e+02	680	6	1.2	1316	1	RPOC_MYCLE	DNA-DIRECTED RNA POLYM	1.29e+02

389	1.2	416	1	CP11_RAT	CONTRAPSIN-LIKE PROTEA	1.29e+02	462	6	1.2	500	1	PTGI_BOVIN	PROSTACYCLIN SYNTHASE	1.29e+02
390	1.2	419	1	SYS_MYCTU	SERYL-TRNA SYNTHETASE	1.29e+02	463	6	1.2	501	1	DHE3_BOVIN	GLUTAMATE DEHYDROGENAS	1.29e+02
391	1.2	420	1	CRFR_CHICK	CORTICOTROPIN RELEASIN	1.29e+02	464	6	1.2	501	1	SCRB_PEDPE	SUCROSE-6-PHOSPHATE HY	1.29e+02
392	1.2	420	1	ML1C_XENIA	MELATONIN RECEPTOR TYP	1.29e+02	465	6	1.2	502	1	Y190_HELPY	HYPOTHETICAL PROTEIN H	1.29e+02
393	1.2	422	1	PAX6_XENIA	PAIRED BOX PROTEIN PAX	1.29e+02	466	6	1.2	503	1	YOW5_CABEL	HYPOTHETICAL 57.1 KD P	1.29e+02
394	1.2	422	1	PAX6_HUMAN	PAIRED BOX PROTEIN PAX	1.29e+02	467	6	1.2	503	1	DHE3_CHICK	GLUTAMATE DEHYDROGENAS	1.29e+02
395	1.2	422	1	PAX6_MOUSE	PAIRED BOX PROTEIN PAX	1.29e+02	468	6	1.2	505	1	MATK_SAXIN	PROBABLE INTRON MATURA	1.29e+02
396	1.2	422	1	PAX6_RAT	PAIRED BOX PROTEIN PAX	1.29e+02	469	6	1.2	506	1	KPY2_YEAST	TYROSINE KINASE 2 (EC	1.29e+02
397	1.2	423	1	YA95_METJA	HYPOTHETICAL PROTEIN M	1.29e+02	470	6	1.2	508	1	TY3H_DROME	ATP SYNTHASE ALPHA CHA	1.29e+02
398	1.2	423	1	EF1A_ARCTU	ELONGATION FACTOR 1-AL	1.29e+02	471	6	1.2	509	1	VP67_MAIZE	MAJOR ENVELOPE GLYCOPR	1.29e+02
399	1.2	423	1	IRK6_HUMAN	G PROTEIN-ACTIVATED IN	1.29e+02	472	6	1.2	509	1	VE2_HPV36	REGULATORY PROTEIN E2.	1.29e+02
400	1.2	424	1	ZP3A_CALSQ	ZONA PELLUCIDA SPERM-B	1.29e+02	473	6	1.2	510	1	YJCW_ECOLI	HYPOTHETICAL ABC TRANS	1.29e+02
401	1.2	425	1	IRK6_MOUSE	G PROTEIN-ACTIVATED IN	1.29e+02	474	6	1.2	512	1	VP67_NPVAC	MAJOR ENVELOPE GLYCOPR	1.29e+02
402	1.2	425	1	IRK6_RAT	G PROTEIN-ACTIVATED IN	1.29e+02	475	6	1.2	513	1	SYS_METHH	SERYL-TRNA SYNTHETASE	1.29e+02
403	1.2	425	1	IRK6_MESAU	G PROTEIN-ACTIVATED IN	1.29e+02	476	6	1.2	514	1	ALAB_MOUSE	ALPHA-1B ADRENERGIC RE	1.29e+02
404	1.2	426	1	WCAC_ECOLI	COLANIC ACID BIOSYNTHET	1.29e+02	477	6	1.2	515	1	ENV_BLVAV	ENV POLYPROTEIN PRECUR	1.29e+02
405	1.2	427	1	GBP2_YEAST	SINGLE-STRAND TELOMERI	1.29e+02	478	6	1.2	515	1	ENV_BLV	ENV POLYPROTEIN PRECUR	1.29e+02
406	1.2	428	1	SYS_STAUF	SERYL-TRNA SYNTHETASE	1.29e+02	479	6	1.2	515	1	ENV_BLVAV	ENV POLYPROTEIN PRECUR	1.29e+02
407	1.2	433	1	LE21_ARCTU	PUTATIVE 3-ISOPROPYLMA	1.29e+02	480	6	1.2	515	1	ALAB_MESAU	ALPHA-1B ADRENERGIC RE	1.29e+02
408	1.2	435	1	YCW9_YEAST	HYPOTHETICAL 48.5 KD P	1.29e+02	481	6	1.2	515	1	ENV_BLV	ENV POLYPROTEIN PRECUR	1.29e+02
409	1.2	441	1	VE2_HPV38	REGULATORY PROTEIN E2.	1.29e+02	482	6	1.2	515	1	ENV_BLV	ENV POLYPROTEIN PRECUR	1.29e+02
410	1.2	442	1	HRDB_STRCO	RNA POLYMERASE PRINCIP	1.29e+02	483	6	1.2	515	1	ENV_BLVAV	ENV POLYPROTEIN PRECUR	1.29e+02
411	1.2	444	1	CRFR_HUMAN	CORTICOTROPIN RELEASIN	1.29e+02	484	6	1.2	515	1	ENV_BLVAV	ENV POLYPROTEIN PRECUR	1.29e+02
412	1.2	445	1	BPI_RABIT	BACTERICIDAL PERMEABIL	1.29e+02	485	6	1.2	515	1	ALAB_RAT	ALPHA-1B ADRENERGIC RE	1.29e+02
413	1.2	448	1	OCT6_HUMAN	OCTAMER-BINDING TRANSC	1.29e+02	486	6	1.2	515	1	VAC1_YEAST	VAC1 PROTEIN.	1.29e+02
414	1.2	449	1	OCT6_MOUSE	OCTAMER-BINDING TRANSC	1.29e+02	487	6	1.2	516	1	YB4F_SCHPO	HYPOTHETICAL 57.6 KD P	1.29e+02
415	1.2	449	1	LIPP_PIG	TRIACYLGLYCEROL LIPASE	1.29e+02	488	6	1.2	519	1	KPYR_CANFA	PYRUVATE KINASE, ISOZY	1.29e+02
416	1.2	451	1	OCT6_RAT	OCTAMER-BINDING TRANSC	1.29e+02	489	6	1.2	519	1	ALAB_HUMAN	ALPHA-1B ADRENERGIC RE	1.29e+02
417	1.2	456	1	YDE8_SCHPO	HYPOTHETICAL 52.4 KD P	1.29e+02	490	6	1.2	521	1	Y029_MYCTU	HYPOTHETICAL 56.0 KD P	1.29e+02
418	1.2	460	1	YX16_MYCTU	HYPOTHETICAL 51.3 KD P	1.29e+02	491	6	1.2	523	1	YHUS_ECOLI	HYPOTHETICAL 59.4 KD P	1.29e+02
419	1.2	460	1	SUOX_CHICK	SULFITE OXIDASE (EC 1.	1.29e+02	492	6	1.2	527	1	DPE2_HUMAN	DNA POLYMERASE EPSILON	1.29e+02
420	1.2	461	1	VE2_HPV09	REGULATORY PROTEIN E2.	1.29e+02	493	6	1.2	528	1	AK_YEAST	ASPARTOKINASE (EC 2.7.	1.29e+02
421	1.2	463	1	ARLY_SACDO	ARGININOSUCCINATE LYAS	1.29e+02	494	6	1.2	528	1	YCD8_YEAST	HYPOTHETICAL 58.8 KD P	1.29e+02
422	1.2	463	1	ARLY_YEAST	ARGININOSUCCINATE LYAS	1.29e+02	495	6	1.2	530	1	PRO9_YEAST	PRE-MRNA SPLICING FACT	1.29e+02
423	1.2	465	1	LIPP_RABIT	TRIACYLGLYCEROL LIPASE	1.29e+02	496	6	1.2	531	1	FMO1_PIG	DIMETHYLANILINE MONOOX	1.29e+02
424	1.2	466	1	CYP8_CAEEL	PEPTIDYL-PROLYL CIS-TR	1.29e+02	497	6	1.2	532	1	FMO1_RAT	DIMETHYLANILINE MONOOX	1.29e+02
425	1.2	466	1	ALAA_RAT	ALPHA-1A ADRENERGIC RE	1.29e+02	498	6	1.2	532	1	FMO1_MOUSE	DIMETHYLANILINE MONOOX	1.29e+02
426	1.2	466	1	ALAA_RABIT	ALPHA-1A ADRENERGIC RE	1.29e+02	499	6	1.2	534	1	YAJE_SCHPO	HYPOTHETICAL 59.0 KD P	1.29e+02
427	1.2	466	1	ALAA_HUMAN	ALPHA-1A ADRENERGIC RE	1.29e+02	500	6	1.2	534	1	FMO1_RABIT	DIMETHYLANILINE MONOOX	1.29e+02
428	1.2	466	1	ALAA_BOVIN	ALPHA-1A ADRENERGIC RE	1.29e+02	501	6	1.2	542	1	RESB_RABIT	DIMETHYLANILINE MONOOX	1.29e+02
429	1.2	467	1	RFAL_CRIFA	REPLICATION FACTOR A 5	1.29e+02	502	6	1.2	547	1	FIB1_ADE40	RESB PROTEIN.	1.29e+02
430	1.2	467	1	ATPX_BACFI	ATP SYNTHASE BETA CHAI	1.29e+02	503	6	1.2	549	1	FIB1_ADE40	FIBER PROTEIN 1.	1.29e+02
431	1.2	468	1	NAH_SCHPO	NA(+)/H(+) ANTIPTOTER.	1.29e+02	504	6	1.2	553	1	TREF_ECOLI	PROBABLE CYTOPLASMIC T	1.29e+02
432	1.2	469	1	ATPB_STRMU	ATP SYNTHASE BETA CHAI	1.29e+02	505	6	1.2	554	1	PPCK_CANAL	PHOSPHOENOLPYRUVATE CA	1.29e+02
433	1.2	470	1	YICO_ECOLI	HYPOTHETICAL 49.9 KD P	1.29e+02	506	6	1.2	554	1	NBL4_MOUSE	3-(3-HYDROXY-PHENYL)PR	1.29e+02
434	1.2	470	1	ALAA_ORYLA	ALPHA-1A ADRENERGIC RE	1.29e+02	507	6	1.2	557	1	YAH6_SCHPO	HYPOTHETICAL 62.7 KD P	1.29e+02
435	1.2	471	1	PPAV_CAEEL	PUTATIVE ACID PHOSPHAT	1.29e+02	508	6	1.2	558	1	DHE3_MOUSE	GLUTAMATE DEHYDROGENAS	1.29e+02
436	1.2	471	1	ATPB_BACME	ATP SYNTHASE BETA CHAI	1.29e+02	509	6	1.2	558	1	DHE3_HUMAN	GLUTAMATE DEHYDROGENAS	1.29e+02
437	1.2	472	1	NIFE_AZOVI	NITROGENASE IRON-MOLYB	1.29e+02	510	6	1.2	562	1	DHE3_RAT	GLUTAMATE DEHYDROGENAS	1.29e+02
438	1.2	474	1	U2AF_HUMAN	SPLICING FACTOR U2AF 6	1.29e+02	511	6	1.2	564	1	FIB1_ADE41	FIBER PROTEIN 1.	1.29e+02
439	1.2	475	1	U2AF_MOUSE	SPLICING FACTOR U2AF 6	1.29e+02	512	6	1.2	564	1	HSE2_CHICK	HEAT SHOCK FACTOR PROT	1.29e+02
440	1.2	475	1	U2AF_MOUSE	SPLICING FACTOR U2AF 6	1.29e+02	513	6	1.2	566	1	HEMA_IATIN	HEMAGGLUTININ PRECURSO	1.29e+02
441	1.2	476	1	BM3B_RAT	BONE MORPHOGENETIC PRO	1.29e+02	514	6	1.2	566	1	HEMA_IATIN	HEMAGGLUTININ PRECURSO	1.29e+02
442	1.2	476	1	BM3B_MOUSE	BONE MORPHOGENETIC PRO	1.29e+02	515	6	1.2	566	1	HEMA_IATIN	HEMAGGLUTININ PRECURSO	1.29e+02
443	1.2	477	1	OAR_HELVI	OCTOPAMINE RECEPTOR.	1.29e+02	516	6	1.2	566	1	HEMA_IATIN	HEMAGGLUTININ PRECURSO	1.29e+02
444	1.2	478	1	BM3B_HUMAN	BONE MORPHOGENETIC PRO	1.29e+02	517	6	1.2	566	1	HEMA_IATIN	HEMAGGLUTININ PRECURSO	1.29e+02
445	1.2	479	1	VGIC_HSV2G	GLYCOPROTEIN C PRECURS	1.29e+02	518	6	1.2	566	1	HEMA_IATIN	HEMAGGLUTININ PRECURSO	1.29e+02
446	1.2	479	1	OAR_BOMMO	OCTOPAMINE RECEPTOR.	1.29e+02	519	6	1.2	566	1	HEMA_IATIN	HEMAGGLUTININ PRECURSO	1.29e+02
447	1.2	479	1	ICE8_HUMAN	CASPASE-8 PRECURSOR (E	1.29e+02	520	6	1.2	568	1	HEMA_IAPUE	HEMAGGLUTININ PRECURSO	1.29e+02
448	1.2	480	1	VGIC_HSV23	GLYCOPROTEIN C PRECURS	1.29e+02	521	6	1.2	568	1	GGT_MOUSE	GAMMA-GLUTAMYLTRANSPER	1.29e+02
449	1.2	480	1	VGIC_HSV2H	GLYCOPROTEIN C PRECURS	1.29e+02	522	6	1.2	569	1	GGT_MOUSE	GAMMA-GLUTAMYLTRANSPER	1.29e+02
450	1.2	481	1	LMRA_STRLN	LINCOMYCIN RESISTANCE	1.29e+02	523	6	1.2	569	1	YAR3_SCHPO	HYPOTHETICAL 61.1 KD P	1.29e+02
451	1.2	482	1	NF31_NAEFO	VIRULENCE-RELATED PROT	1.29e+02	524	6	1.2	570	1	NODU_BRAVA	MODULATION PROTEIN U (1.29e+02
452	1.2	484	1	PEPD_ECOLI	AMINOACYL-HISTIDINE DI	1.29e+02	525	6	1.2	570	1	SID1_USTAE	L-ORNITHINE 5-MONOOXYG	1.29e+02
453	1.2	490	1	CLK3_MOUSE	PROTEIN KINASE CLK3 (E	1.29e+02	526	6	1.2	574	1	KPYR_HUMAN	PYRUVATE KINASE, ISOZY	1.29e+02
454	1.2	490	1	CLK3_HUMAN	PROTEIN KINASE CLK3 (E	1.29e+02	527	6	1.2	574	1	KPYR_HUMAN	PYRUVATE KINASE, ISOZY	1.29e+02
455	1.2	490	1	BCH2_RHOSH	CHLOROPHYLLIDE REDUCTA	1.29e+02	528	6	1.2	574	1	KPYR_MOUSE	PYRUVATE KINASE, ISOZY	1.29e+02
456	1.2	490	1	CLK3_RAT	PROTEIN KINASE CLK3 (E	1.29e+02	529	6	1.2	574	1	KPYR_MOUSE	PYRUVATE KINASE, ISOZY	1.29e+02
457	1.2	492	1	HUPR_RHOCA	HYDROGENASE TRANSCRIPT	1.29e+02	530	6	1.2	578	1	KPK7_ARATH	PUTATIVE SERINE/THREON	1.29e+02
458	1.2	494	1	PHYS_AVESE	PHYTOCHROME A TYPE 5 (1.29e+02	531	6	1.2	583	1	YCV1_YEAST	PUTATIVE SERINE/THREON	1.29e+02
459	1.2	496	1	IAP3_MOUSE	INHIBITOR OF APOPTOSIS	1.29e+02	532	6	1.2	590	1	FTSZ_AGRTU	CELL DIVISION PROTEIN	1.29e+02
460	1.2	498	1	PSG4_MOUSE	PREGNANCY-SPECIFIC BET	1.29e+02	533	6	1.2	592	1	FTSZ_RHIME	CELL DIVISION PROTEIN	1.29e+02
461	1.2	500	1	PTGI_HUMAN	PROSTACYCLIN SYNTHASE	1.29e+02	534	6	1.2	592	1	FTSZ_BARBA	CELL DIVISION PROTEIN	1.29e+02

243	6	1.2	269	1	COX3_TRIRU	CYTOCHROME C OXIDASE P	1.29e+02	316	6	1.2	355	1	OPSG_CHICK	GREEN-SENSITIVE OPSIN	1.29e+02
244	6	1.2	271	1	YABG_RHISN	HYPOTHETICAL PROTEIN Y	1.29e+02	317	6	1.2	355	1	OPSB_GEGGE	BLUE-SENSITIVE OPSIN P	1.29e+02
245	6	1.2	271	1	ECEL_CANAL	CELL ELONGATION PROTEI	1.29e+02	318	6	1.2	355	1	OPSB_ANOCA	BLUE-SENSITIVE OPSIN (1.29e+02
246	6	1.2	274	1	XAP5_HUMAN	XAP-5 PROTEIN (FRAGMENT	1.29e+02	319	6	1.2	355	1	CM4T_STRPE	CARMINOMYCIN 4-O-METHY	1.29e+02
247	6	1.2	275	1	SRP1_SCHPO	SRP1 PROTEIN.	1.29e+02	320	6	1.2	356	1	RS41_ARATH	ARGININE/SERINE-RICH S	1.29e+02
248	6	1.2	277	1	ICE3_CRIL0	APOPAIN PRECURSOR (EC	1.29e+02	321	6	1.2	359	1	OPSG_MOUSE	GREEN-SENSITIVE OPSIN	1.29e+02
249	6	1.2	277	1	ICE3_RAT	APOPAIN PRECURSOR (EC	1.29e+02	322	6	1.2	361	1	OPSB_CHICK	BLUE-SENSITIVE OPSIN (1.29e+02
250	6	1.2	277	1	ARGE_LEPBI	POSSIBLE ACETYLORNTIHI	1.29e+02	323	6	1.2	364	1	GTP1_SCHPO	GTP-BINDING PROTEIN 1.	1.29e+02
251	6	1.2	277	1	ICE3_HUMAN	APOPAIN PRECURSOR (EC	1.29e+02	324	6	1.2	364	1	OPSR_CAPHI	RED-SENSITIVE OPSIN (R	1.29e+02
252	6	1.2	277	1	ICE3_MOUSE	APOPAIN PRECURSOR (EC	1.29e+02	325	6	1.2	364	1	OPSG_HUMAN	RED-SENSITIVE OPSIN (R	1.29e+02
253	6	1.2	281	1	TFE1_MOUSE	TRANSCRIPTION FACTOR E	1.29e+02	326	6	1.2	364	1	OPSG_HUMAN	GREEN-SENSITIVE OPSIN	1.29e+02
254	6	1.2	283	1	YB89_METJA	HYPOTHETICAL PROTEIN M	1.29e+02	327	6	1.2	364	1	OPSG_RABIT	GREEN-SENSITIVE OPSIN	1.29e+02
255	6	1.2	283	1	DAA4_BACSH	D-ALANINE AMINOTRANSFE	1.29e+02	328	6	1.2	365	1	YAGB_SCHPO	HYPOTHETICAL 4.2.3 KD P	1.29e+02
256	6	1.2	285	1	ATPG_BACME	ATP SYNTHASE GAMMA CHA	1.29e+02	329	6	1.2	366	1	DP3B_SALTY	DNA POLYMERASE III, BE	1.29e+02
257	6	1.2	286	1	VNS1_INCCA	NONSTRUCTURAL PROTEIN	1.29e+02	330	6	1.2	366	1	DP3B_SCHPO	DNA POLYMERASE III, BE	1.29e+02
258	6	1.2	288	1	CBAB_ALCSP	3-CHLOROBENZOATE-3,4-D	1.29e+02	331	6	1.2	367	1	MTCL_CHV13	MODIFICATION METHYLASE	1.29e+02
259	6	1.2	288	1	MOVF_PSVJ	CELL-TO-CELL MOVEMENT	1.29e+02	332	6	1.2	368	1	ZP50_BRARE	POU DOMAIN PROTEIN ZP-	1.29e+02
260	6	1.2	289	1	YFK2_YEAST	HYPOTHETICAL 31.9 KD P	1.29e+02	333	6	1.2	372	1	ALAT_CYPCA	ALPHA-1-ANTITRYPSIN HO	1.29e+02
261	6	1.2	289	1	MT61_METJA	PROBABLE MODIFICATION	1.29e+02	334	6	1.2	372	1	LEM1_MOUSE	L-SELECTIN PRECURSOR (1.29e+02
262	6	1.2	289	1	ATPG_KLULA	ATP SYNTHASE GAMMA CHA	1.29e+02	335	6	1.2	373	1	YFJ4_YEAST	HYPOTHETICAL 40.4 KD P	1.29e+02
263	6	1.2	290	1	SSO1_YEAST	SSO1 PROTEIN.	1.29e+02	336	6	1.2	373	1	ICEB_MOUSE	CASPASE-11 PRECURSOR (1.29e+02
264	6	1.2	290	1	HTRL_ECOLI	HTRL PROTEIN.	1.29e+02	337	6	1.2	374	1	VLEF1_NPVOP	VERY LATE EXPRESSION F	1.29e+02
265	6	1.2	294	1	DRAG_RHORU	ADP-RIBOSYLGLYCERYL	1.29e+02	338	6	1.2	376	1	YREC_VIBCH	HYPOTHETICAL 41.3 KD P	1.29e+02
266	6	1.2	296	1	Z124_HUMAN	ZINC FINGER PROTEIN 12	1.29e+02	339	6	1.2	376	1	YB11_YEAST	HYPOTHETICAL 44.4 KD P	1.29e+02
267	6	1.2	296	1	PERX_BRARA	PEROXIDASE P7 (EC 1.11	1.29e+02	340	6	1.2	377	1	ICE4_HUMAN	CASPASE-4 PRECURSOR (E	1.29e+02
268	6	1.2	297	1	Y727_METJA	MAJOR SURFACE GLYCOPRO	1.29e+02	341	6	1.2	378	1	MOBL_THIFE	MOBILIZATION PROTEIN M	1.29e+02
269	6	1.2	298	1	VGLG_HRSVA	MAJOR SURFACE GLYCOPRO	1.29e+02	342	6	1.2	381	1	GAL1_ECOLI	GALACTOKINASE (EC 2.7.	1.29e+02
270	6	1.2	298	1	VGLG_HRSVA	MAJOR SURFACE GLYCOPRO	1.29e+02	343	6	1.2	381	1	CYB_SARHA	CYTOCHROME B (EC 1.10.	1.29e+02
271	6	1.2	298	1	VGLG_HRSVA	MAJOR SURFACE GLYCOPRO	1.29e+02	344	6	1.2	381	1	THIT_METJA	PROBABLE THIAMIN BIOSY	1.29e+02
272	6	1.2	298	1	FEOB_SALTY	FERROUS IRON TRANSPORT	1.29e+02	345	6	1.2	381	1	CYB_DASHA	CYTOCHROME B (EC 1.10.	1.29e+02
273	6	1.2	301	1	MCP_BPF41	MAJOR CAPSID PROTEIN (E	1.29e+02	346	6	1.2	382	1	NCAP_CVPPU	NUCLEOCAPSID PROTEIN (1.29e+02
274	6	1.2	303	1	ICE7_MOUSE	CASPASE-7 PRECURSOR (E	1.29e+02	347	6	1.2	382	1	NCAP_CVPP8	NUCLEOCAPSID PROTEIN (1.29e+02
275	6	1.2	304	1	THTR_YEAST	POTATIVE THIOSULFATE S	1.29e+02	348	6	1.2	382	1	NCAP_CVPP8	NUCLEOCAPSID PROTEIN (1.29e+02
276	6	1.2	306	1	KDGR_ERWCH	PECTIN DEGRADATION REP	1.29e+02	349	6	1.2	382	1	ICEB_XENLA	INTERLEUKIN-1 BETA CON	1.29e+02
277	6	1.2	309	1	WAP9_COCIM	WALL-ASSOCIATED PROTEI	1.29e+02	350	6	1.2	382	1	CYB_MONDO	CYTOCHROME B (EC 1.10.	1.29e+02
278	6	1.2	309	1	CYP9_CAEEL	PEPTIDYL-PROLYL CIS-TR	1.29e+02	351	6	1.2	382	1	NCAP_CVPRM	NUCLEOCAPSID PROTEIN (1.29e+02
279	6	1.2	311	1	SUFI_HAEIN	SUFI PROTEIN HOMOLOG P	1.29e+02	352	6	1.2	383	1	VE2_HPVI5	REGULATORY PROTEIN E2.	1.29e+02
280	6	1.2	312	1	ATO_DROME	ATONAL PROTEIN.	1.29e+02	353	6	1.2	383	1	GYRA_MYGA	DNA GYRASE SUBUNIT A (1.29e+02
281	6	1.2	312	1	PDXK_HUMAN	PYRIDOXINE KINASE (EC	1.29e+02	354	6	1.2	385	1	KUP1_CAEEL	KUP-1 PROTEIN.	1.29e+02
282	6	1.2	314	1	NIKB_ECOLI	NICKEL TRANSPORT SYSTE	1.29e+02	355	6	1.2	385	1	INTC_ECOLI	PUTATIVE PROPHAGE SF6-	1.29e+02
283	6	1.2	314	1	OLFI_HUMAN	OLFACTORY RECEPTOR-LIK	1.29e+02	356	6	1.2	386	1	SUCC_XENLA	SUCCINYL-COA SYNTHETAS	1.29e+02
284	6	1.2	319	1	COA2_POVM3	COAT PROTEIN VP2 [CONT	1.29e+02	357	6	1.2	386	1	ICEA_XENLA	INTERLEUKIN-1 BETA CON	1.29e+02
285	6	1.2	319	1	COA2_POVMA	COAT PROTEIN VP2 [CONT	1.29e+02	358	6	1.2	388	1	OXYR_RAT	OXYTOCIN RECEPTOR (OT-	1.29e+02
286	6	1.2	319	1	COA2_POVMC	COAT PROTEIN VP2 [CONT	1.29e+02	359	6	1.2	388	1	OXYR_MOUSE	OXYTOCIN RECEPTOR (OT-	1.29e+02
287	6	1.2	323	1	YECF_ECOLI	HYPOTHETICAL 37.0 KD P	1.29e+02	360	6	1.2	388	1	GSPL_ECOLI	PROBABLE GENERAL SECRE	1.29e+02
288	6	1.2	323	1	ICE1_DROME	CASPASE-1 PRECURSOR (E	1.29e+02	361	6	1.2	389	1	Y107_HUMAN	HYPOTHETICAL PROTEIN K	1.29e+02
289	6	1.2	327	1	WZB2_ECOLI	CHAIN LENGTH DETERMINA	1.29e+02	362	6	1.2	393	1	PURT_METJA	PROBABLE PHOSPHORIBOSY	1.29e+02
290	6	1.2	328	1	NIR3_AZOBR	NIFER3-LIKE PROTEIN.	1.29e+02	363	6	1.2	394	1	FTSZ_AZOVI	CELL DIVISION PROTEIN	1.29e+02
291	6	1.2	334	1	API_RAT	TRANSCRIPTION FACTOR A	1.29e+02	364	6	1.2	394	1	FTSZ_PSEAE	CELL DIVISION PROTEIN	1.29e+02
292	6	1.2	334	1	API_MOUSE	TRANSCRIPTION FACTOR A	1.29e+02	365	6	1.2	395	1	YHAC_ECOLI	HYPOTHETICAL 45.2 KD P	1.29e+02
293	6	1.2	336	1	YN66_YEAST	HYPOTHETICAL 39.2 KD P	1.29e+02	366	6	1.2	395	1	NODC_AZOCA	N-ACETYLGLUCOSAMINYLR	1.29e+02
294	6	1.2	339	1	ICE_DROME	CASPASE PRECURSOR (EC	1.29e+02	367	6	1.2	395	1	ILV5_YEAST	KETOL-ACID REDUCTOISOM	1.29e+02
295	6	1.2	339	1	MBR1_YEAST	MBR1 PROTEIN.	1.29e+02	368	6	1.2	397	1	TYRB_ECOLI	AROMATIC-AMINO-ACID AM	1.29e+02
296	6	1.2	340	1	PLDB_ECOLI	LYSOPHOSPHOLIPASE L2 (1.29e+02	369	6	1.2	399	1	BRS3_HUMAN	BOMBESIN RECEPTOR SUBT	1.29e+02
297	6	1.2	341	1	TA2R_RAT	THROMBOXANE A2 RECEPTO	1.29e+02	370	6	1.2	399	1	COAT_NODAV	COAT PROTEIN PRECURSOR	1.29e+02
298	6	1.2	343	1	HRCA_BACSU	HEAT-INDUCIBLE TRANSCR	1.29e+02	371	6	1.2	400	1	FTSZ_PSEPU	CELL DIVISION PROTEIN	1.29e+02
299	6	1.2	344	1	ILVC_LACLA	KETOL-ACID REDUCTOISOM	1.29e+02	372	6	1.2	401	1	YNZ4_YEAST	HYPOTHETICAL 46.2 KD P	1.29e+02
300	6	1.2	344	1	STSY_RAUSE	STRICTOSIDINE SYNTHASE	1.29e+02	373	6	1.2	401	1	RL3_TOXCA	60S RIBOSOMAL PROTEIN	1.29e+02
301	6	1.2	344	1	HEMA_IANAR	HEMAGGLUTININ PRECURSO	1.29e+02	374	6	1.2	402	1	ILBC_MOUSE	INTERLEUKIN-1 BETA CON	1.29e+02
302	6	1.2	344	1	HEMA_IAXIA	HEMAGGLUTININ PRECURSO	1.29e+02	375	6	1.2	402	1	ILBC_MOUSE	INTERLEUKIN-1 BETA CON	1.29e+02
303	6	1.2	344	1	HEMA_IATAT	HEMAGGLUTININ PRECURSO	1.29e+02	376	6	1.2	403	1	HMPA_ALCEU	FLAVOHEMOPROTEIN (HAEM	1.29e+02
304	6	1.2	346	1	YMLC_BACSU	HYPOTHETICAL 37.0 KD P	1.29e+02	377	6	1.2	404	1	GLGC_MYCTU	GLUCOSE-1-PHOSPHATE AD	1.29e+02
305	6	1.2	346	1	Y131_HAEIN	HYPOTHETICAL PROTEIN H	1.29e+02	378	6	1.2	406	1	RTA_RAPTH	HEMOCYANIN (RTA) (FRAG	1.29e+02
306	6	1.2	347	1	OPSV_CHICK	VIOLET-SENSITIVE OPSIN	1.29e+02	379	6	1.2	406	1	IDH_SPHYA	ISOCITRATE DEHYDROGENA	1.29e+02
307	6	1.2	347	1	UTR2_YEAST	UTR2 PROTEIN (UNKNOWN	1.29e+02	380	6	1.2	407	1	YU16_MYCTU	POTATIVE AMINOTRANSFER	1.29e+02
308	6	1.2	349	1	CGM6_HUMAN	CARCINOEMBRYONIC ANTIG	1.29e+02	381	6	1.2	409	1	YNP2_YEAST	HYPOTHETICAL 46.2 KD P	1.29e+02
309	6	1.2	350	1	MUTY_ECOLI	A/G-SPECIFIC ADENINE G	1.29e+02	382	6	1.2	411	1	MSK1_MEDSA	GLYCOCEN SYNTHASE KINA	1.29e+02
310	6	1.2	352	1	WNT2_DROME	PROTEIN DWNT-2 PRECURS	1.29e+02	383	6	1.2	413	1	EFIA_METHH	ELONGATION FACTOR 1-AL	1.29e+02
311	6	1.2	352	1	OPSD_ANGAN	RHODOPSIN, DEEP-SEA FO	1.29e+02	384	6	1.2	415	1	CRFR_MOUSE	CORTICOTROPIN RELEASIN	1.29e+02
312	6	1.2	352	1	OPSF_ANGAN	RHODOPSIN, FRESHWATER	1.29e+02	385	6	1.2	415	1	ERK1_DICDI	EXTRACELLULAR SIGNAL-R	1.29e+02
313	6	1.2	352	1	OPSD_ASTFA	RHODOPSIN.	1.29e+02	386	6	1.2	415	1	CRFR_RAT	CORTICOTROPIN RELEASIN	1.29e+02
314	6	1.2	354	1	HEM6_HUMAN	COPROPORPHYRINOGEN III	1.29e+02	387	6	1.2	415	1	CRFR_SHEEP	CORTICOTROPIN RELEASIN	1.29e+02
315	6	1.2	354	1	HEM6_MOUSE	COPROPORPHYRINOGEN III	1.29e+02	388	6	1.2	416	1	ICE9_HUMAN	CASPASE-9 PRECURSOR (E	1.29e+02

97	7	1.4	488	1	VE2_HPV49	REGULATORY PROTEIN E2.	2.20e+00	170	6	1.2	161	1	SECE_MYCTU	PROBABLE PREPROTEIN TR	1.29e+02
98	7	1.4	493	1	VE2_HPV19	REGULATORY PROTEIN E2.	2.20e+00	171	6	1.2	161	1	ATPX_RHORU	ATP SYNTHASE B' CHAIN	1.29e+02
99	7	1.4	494	1	SFR4_HUMAN	SPLICING FACTOR, ARGIN	2.20e+00	172	6	1.2	162	1	19K_MYCIT	19 KD LIPOPROTEIN ANTI	1.29e+02
100	7	1.4	497	1	VE2_HPV20	REGULATORY PROTEIN E2.	2.20e+00	173	6	1.2	170	1	PLAS_LYCES	PLASTOCYANIN PRECURSOR	1.29e+02
101	7	1.4	498	1	VE2_HPV08	REGULATORY PROTEIN E2.	2.20e+00	174	6	1.2	170	1	HEMX_PROMI	PUTATIVE UROPORPHYRIN-	1.29e+02
102	7	1.4	502	1	VE2_HPV25	REGULATORY PROTEIN E2.	2.20e+00	175	6	1.2	175	1	HEM3_MOUSE	TRANSCRIPTION FACTOR H	1.29e+02
103	7	1.4	503	1	VE2_HPV21	REGULATORY PROTEIN E2.	2.20e+00	176	6	1.2	175	1	HES3_RAT	TRANSCRIPTION FACTOR H	1.29e+02
104	7	1.4	514	1	VE2_HPV5B	REGULATORY PROTEIN E2.	2.20e+00	177	6	1.2	177	1	ATPD_CYACA	ATP SYNTHASE DELTA CHA	1.29e+02
105	7	1.4	514	1	VE2_HPV05	REGULATORY PROTEIN E2.	2.20e+00	178	6	1.2	178	1	YQAC_BACSU	HYPOTHETICAL 20.7 KD P	1.29e+02
106	7	1.4	516	1	CP23_HORVU	SERINE CARBOXYPEPTIDAS	2.20e+00	179	6	1.2	178	1	TRSF_DROER	FEMALE-SPECIFIC TRANSF	1.29e+02
107	7	1.4	615	1	LBR_HUMAN	LAMIN B RECEPTOR (INTE	2.20e+00	180	6	1.2	180	1	Y271_METJA	HYPOTHETICAL PROTEIN M	1.29e+02
108	7	1.4	620	1	Y870_METJA	HYPOTHETICAL PROTEIN M	2.20e+00	181	6	1.2	181	1	CHMU_ERWHE	MONOFUNCTIONAL CHORISM	1.29e+02
109	7	1.4	637	1	LBR_CHICK	LAMIN B RECEPTOR.	2.20e+00	182	6	1.2	184	1	YH01_YEAST	HYPOTHETICAL 21.0 KD P	1.29e+02
110	7	1.4	646	1	SR72_YEAST	SIGNAL RECOGNITION PAR	2.20e+00	183	6	1.2	184	1	TRSF_DROSI	FEMALE-SPECIFIC TRANSF	1.29e+02
111	7	1.4	692	1	YK06_YEAST	HYPOTHETICAL 75.5 KD P	2.20e+00	184	6	1.2	185	1	NUSG_TREPA	TRANSCRIPTION ANTI TERM	1.29e+02
112	7	1.4	715	1	MX2_HUMAN	INTERFERON-REGULATED R	2.20e+00	185	6	1.2	186	1	Y073_BORBU	HYPOTHETICAL PROTEIN B	1.29e+02
113	7	1.4	739	1	DD13_CAEEL	PUTATIVE PRE-MRNA SPLI	2.20e+00	186	6	1.2	187	1	RL18_RAT	60S RIBOSOMAL PROTEIN	1.29e+02
114	7	1.4	759	1	YEN1_YEAST	HYPOTHETICAL 87.4 KD P	2.20e+00	187	6	1.2	187	1	RL18_HUMAN	60S RIBOSOMAL PROTEIN	1.29e+02
115	7	1.4	769	1	CIKE_MOUSE	VOLTAGE-GATED POTASSIU	2.20e+00	188	6	1.2	187	1	PRDB_ECOLI	PREPILIN PEPTIDASE DEP	1.29e+02
116	7	1.4	814	1	SLAP_BACAN	S-LAYER PROTEIN PRECUR	2.20e+00	189	6	1.2	195	1	MTLR_ECOLI	MANNITOL OPERON REPRES	1.29e+02
117	7	1.4	901	1	SOKI_YEAST	SOKI PROTEIN.	2.20e+00	190	6	1.2	195	1	FUMB_METJA	PUTATIVE FUMARATE HYDR	1.29e+02
118	7	1.4	951	1	SFR8_HUMAN	SPLICING FACTOR, ARGIN	2.20e+00	191	6	1.2	196	1	UIDR_ECOLI	UID OPERON REPRESSOR (1.29e+02
119	7	1.4	1070	1	PVDG_PLAKN	DUFFY RECEPTOR, GAMMA	2.20e+00	192	6	1.2	196	1	SFR2_CAEEL	PUTATIVE SPLICING FACT	1.29e+02
120	7	1.4	1150	1	IRRI_YEAST	IRRI PROTEIN.	2.20e+00	193	6	1.2	197	1	TRSF_DROME	FEMALE-SPECIFIC TRANSF	1.29e+02
121	7	1.4	1200	1	DDX8_CAEEL	PUTATIVE PRE-MRNA SPLI	2.20e+00	194	6	1.2	197	1	VCO7_ADEM1	MAJOR CORE PROTEIN PRE	1.29e+02
122	7	1.4	1220	1	DDX8_HUMAN	PROBABLE ATP-DEPENDENT	2.20e+00	195	6	1.2	199	1	PVAC_STRLP	PUROMYCIN N-ACETYLTRAN	1.29e+02
123	7	1.4	1244	1	SLA1_YEAST	CYTOSKELETON ASSEMBLY	2.20e+00	196	6	1.2	200	1	RR4_GUTH	CHLOROPLAST 30S RIBOSO	1.29e+02
124	7	1.4	1339	1	ERB3_RAT	ERBB-3 RECEPTOR PROTEI	2.20e+00	197	6	1.2	203	1	YG3R_YEAST	HYPOTHETICAL 22.2 KD P	1.29e+02
125	7	1.4	1342	1	ERB3_HUMAN	ERBB-3 RECEPTOR PROTEI	2.20e+00	198	6	1.2	205	1	GARI_YEAST	GARI PROTEIN.	1.29e+02
126	7	1.4	1466	1	NKCR_MOUSE	NK-TUMOR RECOGNITION P	2.20e+00	199	6	1.2	208	1	YSX2_CAEEL	HYPOTHETICAL 24.0 KD P	1.29e+02
127	7	1.4	1523	1	SON_HUMAN	SON PROTEIN (SON3).	2.20e+00	200	6	1.2	209	1	YIL3_YEAST	PROBABLE PROTEIN-TYROS	1.29e+02
128	7	1.4	2278	1	FABI_YEAST	PROBABLE PHOSPHATIDYLI	2.20e+00	201	6	1.2	209	1	FTSJ_ECOLI	CELL DIVISION PROTEIN	1.29e+02
129	6	1.2	35	1	PHI1_MYCA	SPERM-SPECIFIC PROTEIN	1.29e+02	202	6	1.2	212	1	ATP6_TROHI	ATP SYNTHASE A CHAIN (1.29e+02
130	6	1.2	37	1	PRT3_SCYCA	PROTAMINE 23 (SCYLLTOR	1.29e+02	203	6	1.2	214	1	VTI1_SCHPO	VESICLE TRANSPORT V-SN	1.29e+02
131	6	1.2	54	1	BVCP_NPVAC	DNA-BINDING PROTEIN (A	1.29e+02	204	6	1.2	214	1	BT33_HUMAN	TRANSCRIPTION FACTOR B	1.29e+02
132	6	1.2	64	1	BVCP_NPVBM	DNA-BINDING PROTEIN (A	1.29e+02	205	6	1.2	214	1	RADC_RHOCA	DNA REPAIR PROTEIN RAD	1.29e+02
133	6	1.2	68	1	YHNV_SALTY	HYPOTHETICAL PROTEIN I	1.29e+02	206	6	1.2	216	1	OLID_HUMAN	OLFACTORY RECEPTOR-LIK	1.29e+02
134	6	1.2	68	1	HSP1_TACAC	SPERM PROTAMINE P1.	1.29e+02	207	6	1.2	216	1	OLIA_HUMAN	OLFACTORY RECEPTOR-LIK	1.29e+02
135	6	1.2	75	1	GVA1_HALSA	GAS VESICLE PROTEIN A.	1.29e+02	208	6	1.2	216	1	VIF_HV2D2	VIRION INFECTIVITY FAC	1.29e+02
136	6	1.2	83	1	YKGI_ECOLI	HYPOTHETICAL 9.1 KD PR	1.29e+02	209	6	1.2	216	1	LUXM_VIBHA	LUXM PROTEIN.	1.29e+02
137	6	1.2	85	1	RPOD_CYACA	DNA-DIRECTED RNA POLYM	1.29e+02	210	6	1.2	219	1	REP2_SCHPO	TRANSCRIPTIONAL ACTIVA	1.29e+02
138	6	1.2	87	1	MAT8_HUMAN	CHLORIDE CONDUCTANCE I	1.29e+02	211	6	1.2	222	1	PHZG_PSEFL	PHENAZINE BIOSYNTHESIS	1.29e+02
139	6	1.2	91	1	YAAD_SCHPO	HYPOTHETICAL 10.5 KD P	1.29e+02	212	6	1.2	225	1	KAD2_YEAST	ADENYLATE KINASE 2 (EC	1.29e+02
140	6	1.2	91	1	PHI1_MYTED	SPERM-SPECIFIC PROTEIN	1.29e+02	213	6	1.2	226	1	ATP6_PONPP	ATP SYNTHASE A CHAIN (1.29e+02
141	6	1.2	95	1	ACP_STRMM	OXYTETRACYCLINE POLYKE	1.29e+02	214	6	1.2	227	1	NODW_BRAJA	NODULATION PROTEIN W.	1.29e+02
142	6	1.2	99	1	PLAS_CAPBU	PLASTOCYANIN.	1.29e+02	215	6	1.2	229	1	CMF3_BACSU	COMP OPERON PROTEIN 3.	1.29e+02
143	6	1.2	99	1	PLAS_SOLTU	PLASTOCYANIN.	1.29e+02	216	6	1.2	229	1	S3AG_BACSU	STAGE III SPOULATION	1.29e+02
144	6	1.2	99	1	IT12_SINAL	TRYPSIN INHIBITOR 2 PR	1.29e+02	217	6	1.2	233	1	VRRP_LAMBD	REPLICATION PROTEIN P.	1.29e+02
145	6	1.2	100	1	YNT8_YEAST	VERY HYPOTHETICAL 11.8	1.29e+02	218	6	1.2	234	1	SCP3_MESAU	SYNAPTONEMAL COMPLEX P	1.29e+02
146	6	1.2	100	1	VG6_BPBO3	EARLY PROTEIN GP6.	1.29e+02	219	6	1.2	237	1	ATPO_ARATH	ATP SYNTHASE DELTA CHA	1.29e+02
147	6	1.2	103	1	RS10_BORBU	30S RIBOSOMAL PROTEIN	1.29e+02	220	6	1.2	238	1	DEOD_HAEIN	PURINE NUCLEOSIDE PHOS	1.29e+02
148	6	1.2	104	1	ARTA_ECOLI	ARTA PROTEIN.	1.29e+02	221	6	1.2	243	1	YJFH_ECOLI	HYPOTHETICAL TRNA/RRNA	1.29e+02
149	6	1.2	106	1	YB91_YEAST	HYPOTHETICAL 12.4 KD P	1.29e+02	222	6	1.2	244	1	TPIS_MYCPN	TRIOSEPHOSPHATE ISOMER	1.29e+02
150	6	1.2	110	1	HIS2_AZOC	PHOSPHORIBOSYL-ATP CYC	1.29e+02	223	6	1.2	246	1	DMG_SQUAC	PROTEOLIPID PROTEIN DM	1.29e+02
151	6	1.2	114	1	YCR2_YEAST	HYPOTHETICAL 13.2 KD P	1.29e+02	224	6	1.2	248	1	SDGF_MOUSE	SCHMANNOMA-DERIVED GRO	1.29e+02
152	6	1.2	117	1	VPS_BPBRD	PROTEIN S (GPS).	1.29e+02	225	6	1.2	253	1	CYS2_ECOLI	CYS2 PROTEIN.	1.29e+02
153	6	1.2	118	1	YNIF_AZOBR	HYPOTHETICAL 12.3 KD P	1.29e+02	226	6	1.2	254	1	4IBL_HUMAN	4-IBB LIGAND (4-IBBL).	1.29e+02
154	6	1.2	120	1	RL7_SALTY	50S RIBOSOMAL PROTEIN	1.29e+02	227	6	1.2	256	1	ADH_ZAPTU	ALCOHOL DEHYDROGENASE	1.29e+02
155	6	1.2	120	1	RL7_ECOLI	50S RIBOSOMAL PROTEIN	1.29e+02	228	6	1.2	257	1	ATP6_SCHPO	ATP SYNTHASE A CHAIN P	1.29e+02
156	6	1.2	122	1	HEMA_IACAO	HEMAGGLUTININ (FRAGMEN	1.29e+02	229	6	1.2	257	1	SCD3_RAT	SYNAPTONEMAL COMPLEX P	1.29e+02
157	6	1.2	125	1	PSAE_SPIOL	PHOTOSYSTEM I REACTION	1.29e+02	230	6	1.2	260	1	MCRG_METVA	HYPOTHETICAL ABC TRANS	1.29e+02
158	6	1.2	131	1	STP2_BOVIN	NUCLEAR TRANSITION PRO	1.29e+02	231	6	1.2	261	1	URRH_HAEIN	METHYL-COENZYME M REDU	1.29e+02
159	6	1.2	133	1	LGUL_VIBPA	PROBABLE LACTOYLGLUTAT	1.29e+02	232	6	1.2	262	1	Y114_METTH	UREASE ACCESSORY PROTE	1.29e+02
160	6	1.2	135	1	LGUL_HAEIN	LACTOYLGLUTATHIONE LYA	1.29e+02	233	6	1.2	262	1	PLC1_CAEEL	HYPOTHETICAL PROTEIN M	1.29e+02
161	6	1.2	138	1	STP2_HUMAN	NUCLEAR TRANSITION PRO	1.29e+02	234	6	1.2	263	1	YXIS_SACER	PUTATIVE 1-ACYL-SN-GLY	1.29e+02
162	6	1.2	139	1	YW37_MYCTU	HYPOTHETICAL 14.7 KD P	1.29e+02	235	6	1.2	264	1	KDGR_ECOLI	HYPOTHETICAL 28.9 KD P	1.29e+02
163	6	1.2	142	1	IGF_MYXGL	INSULIN-LIKE GROWTH FA	1.29e+02	236	6	1.2	264	1	TRA2_DROME	TRANSFORMER-2 SEX-DETE	1.29e+02
164	6	1.2	142	1	RECA_STRCT	RECA PROTEIN (FRAGMENT	1.29e+02	237	6	1.2	265	1	HEA2_ANAVN	HESA PROTEIN, VEGETATI	1.29e+02
165	6	1.2	146	1	YDPI_SULAC	HYPOTHETICAL 16.4 KD P	1.29e+02	238	6	1.2	266	1	PTRC_KLEPN	PTS SYSTEM, SORBOSE-SP	1.29e+02
166	6	1.2	154	1	KMLS_MELGA	TELOKIN.	1.29e+02	239	6	1.2	266	1	CEBD_MOUSE	CCAAT/ENHANCER BINDING	1.29e+02
167	6	1.2	155	1	NDKC_DICDI	NUCLEOSIDE DIPHOSPHATE	1.29e+02	240	6	1.2	268	1	COX3_NEUCR	CYTOCROME C OXIDASE P	1.29e+02
168	6	1.2	157	1	YXJN_BACSU	HYPOTHETICAL 17.4 KD P	1.29e+02	241	6	1.2	269	1	COX3_MAGGR	CYTOCROME C OXIDASE P	1.29e+02
169	6	1.2	158	1	NU6M_MYTED	NADH-UBIQUINONE OXIDOR	1.29e+02	242	6	1.2	269	1			

Query Match 1.4%; Score 7; DB 2; Length 58;
Best Local Similarity 100.0%; Pred. No. 4.44e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 RSRSR 13
QY .150 RSRSR 156

RESULT 15

ENTRY S39425 #type complete
TITLE protamine p1 - duckbill platypus
ORGANISM #formal_name Ornithorhynchus anatinus #common_name duckbill
platypus

DATE 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change
29-Jan-1999

ACCESSIONS
REFERENCE

#authors Retief, J.D.; Winkfein, R.J.; Dixon, G.H.
#journal Eur. J. Biochem. (1993) 218:457-461
#title Evolution of the monotremes. The sequences of the protamine
p1 genes of platypus and echidna.

#cross-references MUID:94094837
#accession S39425 preliminary

GENETICS
#introns 47/1
#molecule_type DNA
#residues 1-61 #label RET
#cross-references EMBL:Z26849

CLASSIFICATION #superfamily sperm histone
SUMMARY #length 61 #molecular_weight 7871 #checksum 4490

Query Match 1.4%; Score 7; DB 2; Length 61;
Best Local Similarity 100.0%; Pred. NO. 4.44e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6 RSRSR 12
QY 150 RSRSR 156

Search completed: Tue Aug 10 11:43:48 1999
Job time : 94 secs.

QY 130 RDSVSSVS 137

RESULT 10

ENTRY 138728 #type complete

TITLE epidermal growth factor receptor kinase substrate - human

ORGANISM #formal_name Homo sapiens #common_name man

DATE 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Jun-1998

ACCESSIONS 138728

REFERENCE 138728

#authors Wong, W.T.; Carlomagno, F.; Druck, T.; Barletta, C.; Croce, C.M.; Huebner, K.; Kraus, M.H.; Di Fiore, P.P.

#journal Oncogene (1994) 9:3057-3061

#title Evolutionary conservation of the EPS8 gene and its mapping to human chromosome 12q23-q24.

#cross-references MUID:94366758

#accession 138728

#status preliminary; translated from GB/EMBL/DBJ

GENETICS #molecule_type mRNA

#residues 1-822 #label RES

#cross-references EMBL:U12535; NID:g530822; PID:g530823

SUMMARY #gene EPS8

#length 822 #molecular-weight 91881 #checksum 5343

Query Match 1.6%; Score 8; DB 2; Length 822;

Best Local Similarity 100.0%; Pred. No. 5.18e-02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 53 RDSVSSVS 60

QY 130 RDSVSSVS 137

RESULT 11

ENTRY S21080 #type fragment

TITLE sperm protein EMI - minor jackknife clam (fragment)

ORGANISM #formal_name Ensis minor #common_name minor jackknife clam

DATE 22-Nov-1993 #sequence_revision 26-Jul-1996 #text_change 23-Feb-1997

ACCESSIONS S21080

REFERENCE S21080

#authors Giancotti, V.; Buratti, E.; Santucci, A.; Neri, P.; Crane-Robinson, C.

#journal Biochim. Biophys. Acta (1992) 1119:296-302

#title Molluscan sperm proteins: Ensis minor.

#cross-references MUID:92190249

#accession S21080

#molecule_type protein

#residues 1-34 #label GIA

KEYWORDS DNA binding; nucleus

SUMMARY #length 34 #checksum 6188

Query Match 1.4%; Score 7; DB 2; Length 34;

Best Local Similarity 100.0%; Pred. No. 4.44e+00;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6 RSRRSR 12

QY 150 RSRRSR 156

RESULT 12

ENTRY C58213 #type complete

TITLE protamine II - American alligator

ORGANISM #formal_name Alligator mississippiensis #common_name American alligator

DATE 08-Nov-1996 #sequence_revision 08-Nov-1996 #text_change 09-May-1997

ACCESSIONS C58213

REFERENCE A58208

#authors Hunt, J.G.; Kasinsky, H.E.; Elsey, R.M.; Wright, C.L.; Rice,

#journal P.; Bell, J.E.; Sharp, D.J.; Kiss, A.J.; Hunt, D.F.; Arnott, D.P.; Russ, M.M.; Shabanowitz, J.; Ausio, J.

#title J. Biol. Chem. (1996) 271:23547-23557

#accession C58213

#status preliminary

#molecule_type protein

#residues 1-56 #label HUN

CLASSIFICATION #superfamily sperm histone

SUMMARY #length 56 #molecular-weight 7537 #checksum 6509

Query Match 1.4%; Score 7; DB 2; Length 56;

Best Local Similarity 100.0%; Pred. No. 4.44e+00;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 RSRRSR 13

QY 150 RSRRSR 156

RESULT 13

ENTRY S34045 #type complete

TITLE protamine - North American opossum

ORGANISM #formal_name Didelphis virginiana, Didelphis marsupialis

DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Sep-1997

ACCESSIONS S34045

REFERENCE S34045

#authors Winkfein, R.J.; Nishikawa, S.; Connor, W.; Dixon, G.H.

#journal Eur. J. Biochem. (1993) 215:63-72

#title Characterization of a marsupial sperm protamine gene and its transcripts from the North American opossum (Didelphis marsupialis).

#cross-references MUID:93345500

#accession S34045

#status preliminary

#molecule_type DNA

#residues 1-58 #label WIN

CLASSIFICATION #superfamily sperm histone

KEYWORDS DNA binding; nucleus

SUMMARY #length 58 #molecular-weight 7941 #checksum 3630

Query Match 1.4%; Score 7; DB 2; Length 58;

Best Local Similarity 100.0%; Pred. No. 4.44e+00;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 RSRRSR 13

QY 150 RSRRSR 156

RESULT 14

ENTRY A58208 #type complete

TITLE protamine I-1 - painted turtle

ORGANISM #formal_name Chrysemys picta #common_name painted turtle

DATE 08-Nov-1996 #sequence_revision 08-Nov-1996 #text_change 09-May-1997

ACCESSIONS A58208

REFERENCE A58208

#authors Hunt, J.G.; Kasinsky, H.E.; Elsey, R.M.; Wright, C.L.; Rice, P.; Bell, J.E.; Sharp, D.J.; Kiss, A.J.; Hunt, D.F.; Arnott, D.P.; Russ, M.M.; Shabanowitz, J.; Ausio, J.

#journal J. Biol. Chem. (1996) 271:23547-23557

#title Protamines of reptiles.

#accession A58208

#status preliminary

#molecule_type protein

#residues 1-58 #label HUN

CLASSIFICATION #superfamily sperm histone

SUMMARY #length 58 #molecular-weight 7620 #checksum 2377

A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Seror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
#cross-references MUID:98044033
#accession A70039
#status preliminary; nucleic acid sequence not shown;
translation not shown

##molecule_type DNA
##residues 1-301 ##label KUN
##cross-references GB:Z99121; GB:AL009126; NID:g2635827; PID:e1186097; PID:g2635922
##experimental_source strain 168

GENETICS
#gene yvfR
CLASSIFICATION #superfamily ATP-binding cassette homology
KEYWORDS P-loop
FEATURE
21-206 #domain ATP-binding cassette homology #label ABC\
38-45 #region nucleotide-binding motif A (P-loop)
SUMMARY #length 301 #molecular-weight 33779 #checksum 1714

Query Match 1.6%; Score 8; DB 2; Length 301;
Best Local Similarity 100.0%; Pred. No. 5.18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 91 LKVDLLE 98
QY 21 LKVDLLE 28

RESULT 7
ENTRY S75895 #type complete
TITLE hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
ORGANISM #formal_name *Synechocystis* sp.
PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998

ACCESSIONS S75895
REFERENCE S74322
#authors Kaneo, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugiura, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
#cross-references MUID:97061201
#accession S75895
#status preliminary
##molecule_type DNA
##residues 1-495 ##label KAN
##cross-references EMBL:D90913; GB:AB001339; NID:g1653348; PID:d1019087; PID:g1653440
##note the nucleotide sequence was submitted to the EMBL Data Library, June 1996
CLASSIFICATION #superfamily xylulokinase

SUMMARY #length 495 #molecular-weight 54351 #checksum 4379

Query Match 1.6%; Score 8; DB 2; Length 495;
Best Local Similarity 100.0%; Pred. No. 5.18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 338 LARSVDSN 345
QY 91 LARSVDSN 98

RESULT 8
ENTRY W2WL47 #type complete
TITLE E2 protein - human papillomavirus type 47
ORGANISM #formal_name human papillomavirus type 47
#note host Homo sapiens (man)
DATE 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 05-Sep-1997

ACCESSIONS D35324
REFERENCE A35324
#authors Kiyono, T.; Adachi, A.; Ishibashi, M.
#journal Virology (1990) 177:401-405
#title Genome organization and taxonomic position of human papillomavirus type 47 inferred from its DNA sequence.
#cross-references MUID:90281611
#accession D35324
#status translation not shown
##molecule_type DNA
##residues 1-506 ##label K1Y
##cross-references GB:M32305; NID:g333062; PID:g333067

CLASSIFICATION #superfamily papillomavirus E2 protein
KEYWORDS DNA binding; early protein; transcription regulation
SUMMARY #length 506 #molecular-weight 57478 #checksum 9705

Query Match 1.6%; Score 8; DB 1; Length 506;
Best Local Similarity 100.0%; Pred. No. 5.18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 302 RARSRSRS 309
QY 148 RARSRSRS 155

RESULT 9
ENTRY S39983 #type complete
TITLE eps8 protein - mouse
ORGANISM #formal_name *Mus musculus* #common_name house mouse
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 17-Mar-1999

ACCESSIONS S39983
REFERENCE S39983
#authors Fazioli, F.; Minichiello, L.; Matoska, V.; Castagnino, P.; Mikl, T.; Wong, W.T.; di Fiore, P.P.
#journal EMBO J. (1993) 12:3799-3808
#title Eps8, a substrate for the epidermal growth factor receptor kinase, enhances EGF-dependent mitogenic signals.
#cross-references MUID:94008987
#accession S39983
#status preliminary
##molecule_type mRNA
##residues 1-821 ##label FAZ
##cross-references EMBL:L21671; NID:g309216; PID:g309217

CLASSIFICATION #superfamily SH3 homology
FEATURE
537-584 #domain SH3 homology #label SH3
SUMMARY #length 821 #molecular-weight 91737 #checksum 9175

Query Match 1.6%; Score 8; DB 2; Length 821;
Best Local Similarity 100.0%; Pred. No. 5.18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 53 RDSVSSVS 60
QY 1111111111

#accession B40082
##molecule_type DNA
##residues 44-159 #label LI2
#accession C40082
##molecule_type protein
##residues 6-15;23-59;93-103;148-170;174-178 #label LI3
REFERENCE A36286
#authors Lin, L.F.H.; Armes, L.G.; Sommer, A.; Smith, D.J.; Collins, F.
#journal J. Biol. Chem. (1990) 265:8942-8947
#title Isolation and characterization of ciliary neurotrophic factor from rabbit sciatic nerves.
#cross-references MUID:90256829
#accession A36286
##molecule_type protein
##residues 6-15;23-59;93-103;140-147;148-169;174-178 #label LI4
COMMENT CNTF has the characteristics of a nonsecreted, cytoplasmic protein; however, a receptor for this protein is found exclusively in nervous and skeletal muscle tissues (see ciliary neurotrophic factor receptor).
COMMENT CNTF promotes survival, neurotransmitter synthesis, and neurite outgrowth in certain neuronal populations.
CLASSIFICATION #superfamily ciliary neurotrophic factor
KEYWORDS growth factor; peripheral neuron
SUMMARY #length 199 #molecular-weight 22662 #checksum 8387
Query Match 1.6%; Score 8; DB 1; Length 199;
Best Local Similarity 100.0%; Pred. No. 5.18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 50 LDSVDGVP 57
QY 370 LDSVDGVP 377
RESULT 4 UNRICE #type complete
ENTRY ciliary neurotrophic factor - rat
TITLE CNTF; survival factor
ALTERNATE_NAMES #formal_name Rattus norvegicus #common_name Norway rat
ORGANISM 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change
DATE 29-May-1998
ACCESSIONS S08144; A43007
REFERENCE S08144
#authors Stoeckli, K.A.; Lottspeich, F.; Sendtner, M.; Masiakowski, P.; Carroll, P.; Goetz, R.; Lindholm, D.; Thoenen, H.
#journal Nature (1989) 342:920-923
#title Molecular cloning, expression and regional distribution of rat ciliary neurotrophic factor.
#cross-references MUID:90081871
#accession S08144
##molecule_type mRNA
##residues 1-200 #label ST1
#cross-references EMBL:X17457; NID:g55968; PID:g55969
#accession A43007
##molecule_type protein
##residues 20-25;29-40;47-60;65-92;113-154;161-166;184-200 #label ST2
COMMENT CNTF has the characteristics of a nonsecreted, cytoplasmic protein; however, a receptor for this protein is found exclusively in nervous and skeletal muscle tissues (see ciliary neurotrophic factor receptor, PIR:I58141).
COMMENT CNTF promotes survival, neurotransmitter synthesis, and neurite outgrowth in certain neuronal populations.
CLASSIFICATION #superfamily ciliary neurotrophic factor
KEYWORDS growth factor; peripheral neuron
SUMMARY #length 200 #molecular-weight 22854 #checksum 2308
Query Match 1.6%; Score 8; DB 1; Length 200;
Best Local Similarity 100.0%; Pred. No. 5.18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 50 LDSVDGVP 57

QY 370 LDSVDGVP 377
RESULT 5
ENTRY JC6125 #type complete
TITLE U2 small nuclear ribonucleoprotein auxiliary factor small chain - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 23-Mar-1997 #sequence_revision 09-May-1997 #text_change
ACCESSIONS JC6125
REFERENCE JC6125
#authors Rudner, D.Z.; Kanaar, R.; Breger, K.S.; Rio, D.C.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:10333-10337
#title Mutations in the small subunit of the Drosophila U2AF splicing factor cause lethality and developmental defects.
#cross-references MUID:96413646
#accession JC6125
##status nucleic acid sequence not shown
##molecule_type mRNA
##residues 1-264 #label RUD
#cross-references GB:U67066; NID:g1621614; PID:g1621615
#experimental_source embryo
COMMENT This protein plays a role in 3' splice site selection. It binds site-specifically to the intron polypyrimidine tract located between the branch site and the 3' splice site of the pre-mRNA and recruits U2 small nuclear ribonucleoprotein to the branch site in the first ATP-dependent step in spliceosome assembly.
GENETICS
#map_position 2 left arm
SUMMARY #length 264 #molecular-weight 29855 #checksum 4187
Query Match 1.6%; Score 8; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 5.18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 193 RARSRSRS 200
QY 148 RARSRSRS 155
RESULT 6
ENTRY A70039 #type complete
TITLE ABC transporter (ATP-binding protein) homolog yvfr - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
ACCESSIONS A70039
REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chol, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrar, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Huilo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Bianchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,

973 5 1.0 558 2 A53719 glutamate dehydrogena 4.86e+03
974 5 1.0 566 1 HMIYV8 hemagglutinin precurs 4.86e+03
975 5 1.0 566 1 HMIYAT hemagglutinin precurs 4.86e+03
976 5 1.0 569 1 KRMSI1 keratin, 59k type I c 4.86e+03
977 5 1.0 613 2 A69535 aldehyde ferredoxin o 4.86e+03
978 5 1.0 614 2 I48385 RNA helicase TN22 - m 4.86e+03
979 5 1.0 622 1 ACCHAN nicotinic acetylcholi 4.86e+03
980 5 1.0 644 2 G70794 probable alpha-isopro 4.86e+03
981 5 1.0 654 2 C69344 heterodisulfide reduc 4.86e+03
982 5 1.0 663 2 S11521 CAMP-gated channel pr 4.86e+03
983 5 1.0 692 2 B70484 conserved hypothetical 4.86e+03
984 5 1.0 747 1 HIBPC7 internal virion prote 4.86e+03
985 5 1.0 770 1 QRHUA4 Alzheimer's disease a 4.86e+03
986 5 1.0 772 2 B65167 hypothetical 88.1 kD 4.86e+03
987 5 1.0 845 1 JDVLRK DNA-directed DNA poly 4.86e+03
988 5 1.0 856 2 F64098 ATP-dependent Clp pro 4.86e+03
989 5 1.0 857 1 D35905 ATP-dependent Clp pro 4.86e+03
990 5 1.0 861 2 B49847 nitrate reductase (EC 4.86e+03
991 5 1.0 908 2 JN0804 nitrate reductase (NA 4.86e+03
992 5 1.0 975 2 S33121 homeotic protein CDP 4.86e+03
993 5 1.0 984 2 JN0658 restriction endonucle 4.86e+03
994 5 1.0 986 1 OYURGA speract receptor prec 4.86e+03
995 5 1.0 1156 2 B70356 chromosome assembly p 4.86e+03
996 5 1.0 1217 1 EGSMG6 epidermal growth fact 4.86e+03
997 5 1.0 1232 2 B70556 probable respiratory 4.86e+03
998 5 1.0 1275 2 S53636 ribulose-bisphosphate 4.86e+03
999 5 1.0 1433 2 S54587 CAT8 protein - yeast 4.86e+03
1000 5 1.0 1447 2 S02160 DNA topoisomerase (AT 4.86e+03

ALIGNMENTS

RESULT 1 A49429 #type complete
ENTRY interleukin-1 beta-converting enzyme homolog CED-3 -
TITLE Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994

ACCESSIONS A49429
REFERENCE A49429
#authors Yuan, J.; Shaham, S.; Ledoux, S.; Ellis, H.M.; Horvitz, H.R.
#journal Cell (1993) 75:641-652
#title The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1 beta-converting enzyme.
#cross-references MUID:94061982
#accession A49429
#status preliminary
#molecule_type DNA
#residues 1-503 #label YUA
#note sequence extracted from NCBI backbone (NCBIN:139825, NCBI:139826)
#length 503 #molecular-weight 56616 #checksum 8929

Query Match 89.9%; Score 452; DB 2; Length 503;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1 MMQRDRSLLENIMFSSHLKVDLELVIAKQVLNSDNGDMINSCGTREKRREIYKA 60
QY 1 MMQRDRSLLENIMFSSHLKVDLELVIAKQVLNSDNGDMINSCGTREKRREIYKA 60
DB 61 VORRGDAFADALRSTGHEGLAEVLEPLARSDVSNAVEFECPMSPASHRRSRALSPA 120
QY 61 VORRGDAFADALRSTGHEGLAEVLEPLARSDVSNAVEFECPMSPASHRRSRALSPA 120
DB 121 GYTSPTRVHRDSYSSVSTSYQDIYSRARSRSRRLHSSDRHNYSSPPVNAFPSPSS 180
QY 121 GYTSPTRVHRDSYSSVSTSYQDIYSRARSRSRRLHSSDRHNYSSPPVNAFPSPSS 180
DB 121 GYTSPTRVHRDSYSSVSTSYQDIYSRARSRSRRLHSSDRHNYSSPPVNAFPSPSS 180
QY 121 GYTSPTRVHRDSYSSVSTSYQDIYSRARSRSRRLHSSDRHNYSSPPVNAFPSPSS 180
DB 181 ANSFTGCSSTLGYSSSRNRSFSKASGPTQYIFHEEDMNFVDAPTISRVEDEKTMYNFSS 240
QY 181 ANSFTGCSSTLGYSSSRNRSFSKASGPTQYIFHEEDMNFVDAPTISRVEDEKTMYNFSS 240

DB 241 PRGMCLINNEHEEQMPTRNGTKADKDNLTNLFRCMGYTVICKDNLTRGMLTIRPAK 300
QY 241 PRGMCLINNEHEEQMPTRNGTKADKDNLTNLFRCMGYTVICKDNLTRGMLTIRPAK 300
DB 301 HESHGDAIVILSHGEENVIGVDIPISTHEIYDLNNAANPRLANKPKIVFQACRG 360
QY 301 HESHGDAIVILSHGEENVIGVDIPISTHEIYDLNNAANPRLANKPKIVFQACRG 360
DB 361 ERRDNGFPVLDSDVGPAPFLRGMNDNDGPLENFLGCVRPQYQVWRKKPSQADILIYA 420
QY 361 ERRDNGFPVLDSDVGPAPFLRGMNDNDGPLENFLGCVRPQYQVWRKKPSQADILIYA 420
DB 421 TTAQYVSWRNSARGSWFIQAVCEVESTHAKDMDVELLTEYAKKVAACGFQTSQGSNLIKQ 480
QY 421 TTAQYVSWRNSARGSWFIQAVCEVESTHAKDMDVELLTEYAKKVAACGFQTSQGSNLIKQ 480
DB 481 MPMTSRLTKRFFWPEARNSAV 503
QY 481 MPMTSRLTKRFFWPEARNSAV 503

RESULT 2 I48723 #type complete
ENTRY ciliary neurotrophic factor - mouse
TITLE #formal_name Mus musculus #common_name house mouse
ORGANISM 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Feb-1997

ACCESSIONS I48723
REFERENCE I48722
#authors Saccone, Y.; Winter, C.G.; Hirsh, D.
#journal Gene (1995) 152:233-238
#title A widely expressed novel C2H2 zinc-finger protein with multiple consensus phosphorylation sites is conserved in mouse and man.
#cross-references MUID:95137394
#accession I48723
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-198 #label RES
#cross-references EMBL:U05342; NID:g453372; PID:g453374
CLASSIFICATION #superfamily ciliary neurotrophic factor
SUMMARY #length 198 #molecular-weight 22587 #checksum 8760

Query Match 1.6%; Score 8; DB 2; Length 198;
Best Local Similarity 100.0%; Pred. No. 5.18e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 50 LDSVDGVP 57
QY 370 LDSVDGVP 377

RESULT 3 UNRBCF #type complete
ENTRY ciliary neurotrophic factor - rabbit
TITLE CNTF; survival factor
ALTERNATE_NAMES #formal_name Oryctolagus cuniculus #common_name domestic rabbit
ORGANISM 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 30-Jun-1993
DATE A40082; B40082; C40082; A36286
ACCESSIONS A40082
REFERENCE A40082
#authors Lin, L.F.H.; Mismar, D.; Lile, J.D.; Armes, L.G.; Butler III, E.T.; Vannice, J.L.; Collins, F.
#journal Science (1989) 246:1023-1025
#title Purification, cloning, and expression of ciliary neurotrophic factor (CNTF).
#cross-references MUID:90069557
#accession A40082
#molecule_type mRNA
#residues 1-199 #label L11
#cross-references GB:M29828

827	6	1.2	1206	2	S72620	probable reverse tran	2.14e+02	900	6	1.2	3084	1	MMMSA	laminin alpha-1 chain	2.14e+02
828	6	1.2	1209	1	DNBEC4	DNA-binding protein -	2.14e+02	901	6	1.2	3144	2	A46068	Huntington disease-as	2.14e+02
829	6	1.2	1219	2	I61713	co-repressor protein	2.14e+02	902	6	1.2	3175	1	RRWVEF	genome polypeptide -	2.14e+02
830	6	1.2	1222	2	S40977	hypothetical protein	2.14e+02	903	6	1.2	3386	1	GNWVDF	genome polypeptide -	2.14e+02
831	6	1.2	1224	2	S73171	DNA-directed RNA poly	2.14e+02	904	6	1.2	3519	2	S43048	polyketide synthase t	2.14e+02
832	6	1.2	1228	2	I40468	surface layer protein	2.14e+02	905	6	1.2	3924	2	S37431	ankyrin 2, neuronal 1	2.14e+02
833	6	1.2	1229	2	A56068	co-repressor protein	2.14e+02	906	6	1.2	3972	2	S75251	hypothetical protein	2.14e+02
834	6	1.2	1232	1	D64413	cobalamin biosynthesi	2.14e+02	907	6	1.2	4391	2	A38096	perlecan precursor -	2.14e+02
835	6	1.2	1254	1	A32686	DNA-directed DNA poly	2.14e+02	908	6	1.2	4427	2	PN0637	polyketide synthase (2.14e+02
836	6	1.2	1259	2	S25954	gene atpa intron 2 pr	2.14e+02	909	6	1.2	4563	1	LPHUB	apolipoprotein B-100	2.14e+02
837	6	1.2	1272	2	S60999	ubiquitin-specific pr	2.14e+02	910	6	1.2	4861	2	S71752	giant protein p619 -	2.14e+02
838	6	1.2	1287	2	S55954	viral mRNA translatio	2.14e+02	911	5	1.0	13	2	B35245	histone H1.c - mouse	4.86e+03
839	6	1.2	1289	1	GUBPT4	proximal tail fiber p	2.14e+02	912	5	1.0	86	2	S35769	T-cell receptor alpha	4.86e+03
840	6	1.2	1290	2	S73653	DNA-directed RNA poly	2.14e+02	913	5	1.0	92	2	S56009	Ig heavy chain variab	4.86e+03
841	6	1.2	1291	2	S02021	microbia polypeptide	2.14e+02	914	5	1.0	108	2	C26405	Ig kappa chain V regi	4.86e+03
842	6	1.2	1292	2	F64237	DNA-directed RNA poly	2.14e+02	915	5	1.0	109	2	S27210	parvalbumin - Japanes	4.86e+03
843	6	1.2	1299	1	WMBEHS	membrane antigen p140	2.14e+02	916	5	1.0	110	1	JN0761	red pigment-concentra	4.86e+03
844	6	1.2	1316	2	S31146	DNA-directed RNA poly	2.14e+02	917	5	1.0	110	1	S57465	Ig lambda chain V-J r	4.86e+03
845	6	1.2	1316	2	G70535	probable rpoC protein	2.14e+02	918	5	1.0	114	1	B31848	calgranulin B - human	4.86e+03
846	6	1.2	1317	2	S77517	DNA-directed RNA poly	2.14e+02	919	5	1.0	120	2	A29350	retrovirus-related po	4.86e+03
847	6	1.2	1323	2	T00037	hypothetical protein	2.14e+02	920	5	1.0	120	2	B24775	T-cell receptor gamma	4.86e+03
848	6	1.2	1332	2	F69732	PBSX prophage ORF xkd	2.14e+02	921	5	1.0	122	2	S40338	Ig kappa chain V regi	4.86e+03
849	6	1.2	1358	2	S64356	phosphoribosylformylg	2.14e+02	922	5	1.0	146	2	S31589	Ig heavy chain V regi	4.86e+03
850	6	1.2	1364	2	T00250	MEGF2 protein - human	2.14e+02	923	5	1.0	146	2	S30434	Ig heavy chain V-D-J	4.86e+03
851	6	1.2	1371	2	S77521	sensory transduction	2.14e+02	924	5	1.0	148	2	S68816	legumin beta chain -	4.86e+03
852	6	1.2	1381	2	S55619	capsid protein 25 - e	2.14e+02	925	5	1.0	153	2	S31700	Ig heavy chain V regi	4.86e+03
853	6	1.2	1386	2	S73401	MG064 homolog R02_orf	2.14e+02	926	5	1.0	159	1	HBBD	hemoglobin beta chain	4.86e+03
854	6	1.2	1396	2	S36851	L-shaped tail fiber p	2.14e+02	927	5	1.0	175	1	CYHUG1	gamma-crystallin 1-2	4.86e+03
855	6	1.2	1402	2	S42748	finger protein - frui	2.14e+02	928	5	1.0	179	2	S55660	capsid protein 65 - e	4.86e+03
856	6	1.2	1403	2	A47328	natural killer cell t	2.14e+02	929	5	1.0	194	2	I39526	hypothetical protein	4.86e+03
857	6	1.2	1405	1	S13421	polymeric globin alph	2.14e+02	930	5	1.0	200	2	G64106	3-isopropylmalate deh	4.86e+03
858	6	1.2	1416	2	D71350	probable DNA-directed	2.14e+02	931	5	1.0	215	2	G49343	probable dibenzothio	4.86e+03
859	6	1.2	1512	2	S72354	genome polypeptide -	2.14e+02	932	5	1.0	217	2	I48902	homeobox protein Pmx	4.86e+03
860	6	1.2	1518	2	S37928	probable purine nucle	2.14e+02	933	5	1.0	228	2	D30857	hypothetical protein	4.86e+03
861	6	1.2	1520	2	T00273	hypothetical protein	2.14e+02	934	5	1.0	237	2	H69820	conserved hypothetical	4.86e+03
862	6	1.2	1536	2	S59841	4-alpha-glucanotransf	2.14e+02	935	5	1.0	239	2	C64188	arginine binding prot	4.86e+03
863	6	1.2	1571	2	T00062	hypothetical protein	2.14e+02	936	5	1.0	245	2	A41858	orfo - Pseudomonas sp	4.86e+03
864	6	1.2	1574	2	G70466	DNA-directed RNA poly	2.14e+02	937	5	1.0	252	2	F65084	hypothetical protein	4.86e+03
865	6	1.2	1592	2	S48933	probable transport pr	2.14e+02	938	5	1.0	253	2	S53614	major prion protein -	4.86e+03
866	6	1.2	1603	2	S17983	gene posterior sex co	2.14e+02	939	5	1.0	266	2	A35037	insulin-like growth f	4.86e+03
867	6	1.2	1676	2	E71410	probable centromere p	2.14e+02	940	5	1.0	267	2	G65103	pts system, n-acetylgl	4.86e+03
868	6	1.2	1676	2	A56508	anucleate primary ste	2.14e+02	941	5	1.0	281	2	A38090	N-hydroxyarylamino O-	4.86e+03
869	6	1.2	1683	2	S56811	probable membrane pro	2.14e+02	942	5	1.0	305	1	PWRYNB	Na+/K+-exchanging ATP	4.86e+03
870	6	1.2	1722	2	I78879	retinoblastoma bindin	2.14e+02	943	5	1.0	365	2	A34840	heterogeneous ribonuc	4.86e+03
871	6	1.2	1729	2	S57596	ribosomal RNA process	2.14e+02	944	5	1.0	369	2	S33603	surfactant protein D	4.86e+03
872	6	1.2	1736	2	A47747	tight junction protei	2.14e+02	945	5	1.0	371	2	JN0450	conglutinin precursor	4.86e+03
873	6	1.2	1742	2	S24600	projectin - fruit fly	2.14e+02	946	5	1.0	372	2	B34261	alanine dehydrogenase	4.86e+03
874	6	1.2	1745	2	S44816	F44E2.1 protein - Cae	2.14e+02	947	5	1.0	379	2	I48133	ubiquinol--cytochrome	4.86e+03
875	6	1.2	1770	2	A71517	hypothetical protein	2.14e+02	948	5	1.0	393	1	A26681	rhizopuspepsin (EC 3.	4.86e+03
876	6	1.2	1786	1	MMHUB1	laminin beta-1 chain	2.14e+02	949	5	1.0	408	2	A69819	probable phosphoester	4.86e+03
877	6	1.2	1805	2	A64224	hypothetical protein	2.14e+02	950	5	1.0	409	2	S12588	pol polypeptide - min	4.86e+03
878	6	1.2	1818	2	S73852	hypothetical protein	2.14e+02	951	5	1.0	409	2	T02776	y4dm protein - Rhizob	4.86e+03
879	6	1.2	1894	2	T02155	DNA-directed DNA poly	2.14e+02	952	5	1.0	412	2	A30605	acyl-CoA dehydrogenas	4.86e+03
880	6	1.2	1922	2	T00637	hypothetical protein	2.14e+02	953	5	1.0	419	1	SYBSYF	tyrosine--tRNA ligase	4.86e+03
881	6	1.2	1927	2	G64585	cag pathogenicity isl	2.14e+02	954	5	1.0	422	2	C70370	dihydroorotase - Agui	4.86e+03
882	6	1.2	1937	2	T03224	probable polypeptide s	2.14e+02	955	5	1.0	449	2	C39926	tailless (tll) protei	4.86e+03
883	6	1.2	2101	2	A42184	NuMA protein - human	2.14e+02	956	5	1.0	450	2	B47265	tailless (tll) protei	4.86e+03
884	6	1.2	2115	2	S23647	NuMA protein - human	2.14e+02	957	5	1.0	458	2	B33560	IL protein - human he	4.86e+03
885	6	1.2	2195	2	S61103	SEC16 protein - yeast	2.14e+02	958	5	1.0	459	2	H64024	hypothetical protein	4.86e+03
886	6	1.2	2252	2	S06188	genome polypeptide 1	2.14e+02	959	5	1.0	468	2	JN0016	peripherin intermedia	4.86e+03
887	6	1.2	2257	2	A57710	acetyl-CoA carboxylas	2.14e+02	960	5	1.0	469	1	NBMS4	C4b-binding protein a	4.86e+03
888	6	1.2	2264	1	GNVVTB	genome polypeptide 1	2.14e+02	961	5	1.0	475	1	Z4BPT4	helicase (EC 3.6.1.-)	4.86e+03
889	6	1.2	2325	2	A61208	chondroitin sulfate p	2.14e+02	962	5	1.0	485	2	D70664	probable gnd protein	4.86e+03
890	6	1.2	2489	2	S59782	probable membrane pro	2.14e+02	963	5	1.0	491	2	S73110	hypothetical protein	4.86e+03
891	6	1.2	2515	2	S47008	tenascin-like protein	2.14e+02	964	5	1.0	504	2	B40829	activin receptor isof	4.86e+03
892	6	1.2	2529	2	A56923	transcription factor	2.14e+02	965	5	1.0	508	1	QTLV6A	photosystem II chloro	4.86e+03
893	6	1.2	2549	2	A54837	rapamycin/FKBP12 targ	2.14e+02	966	5	1.0	514	2	A36793	hypothetical protein	4.86e+03
894	6	1.2	2594	2	S45340	FKBP-rapamycin-associ	2.14e+02	967	5	1.0	521	2	S55318	cytochrome P450 1A, h	4.86e+03
895	6	1.2	2594	2	A35774	kinase-related protei	2.14e+02	968	5	1.0	523	2	D64809	cytochrome d ubiquino	4.86e+03
896	6	1.2	2763	1	WZBE22	gene 22 protein - hum	2.14e+02	969	5	1.0	526	1	KRBOVI	keratin, 54K type I C	4.86e+03
897	6	1.2	2893	2	A64556	toxin-like outer memb	2.14e+02	970	5	1.0	536	2	A40829	activin receptor isof	4.86e+03
898	6	1.2	3063	2	S55505	fatty-acid synthase (2.14e+02	971	5	1.0	545	1	PWBYA	H+-transporting ATP s	4.86e+03
899	6	1.2	3075	2	S14458	laminin alpha-1 chain	2.14e+02	972	5	1.0	555	2	S50157	cyclin-dependent kina	4.86e+03

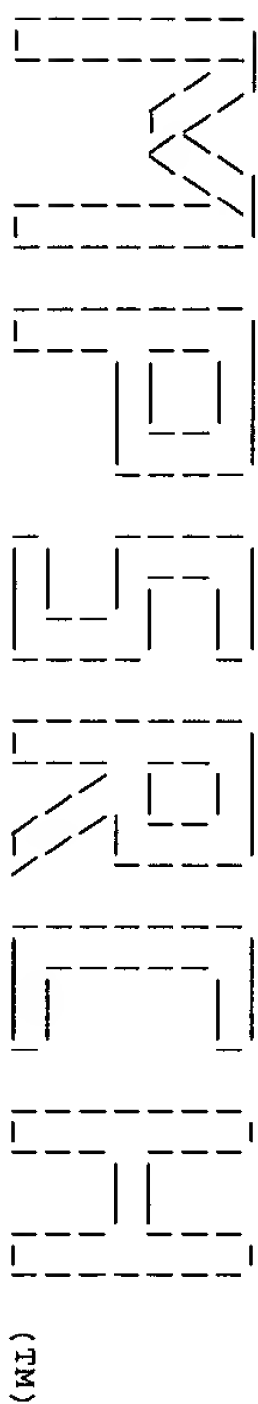
681	6	1.2	638	2	B35816	transcription regulat	2.14e+02	754	6	1.2	814	2	I39627	nicotine dehydrogenas	2.14e+02
682	6	1.2	639	2	H70661	probable dnaG protein	2.14e+02	755	6	1.2	815	2	B56708	extracellular signal-	2.14e+02
683	6	1.2	646	2	G69157	excinuclease ABC chai	2.14e+02	756	6	1.2	816	2	A71006	hypothetical protein	2.14e+02
684	6	1.2	647	2	A37086	beta-galactosidase (E	2.14e+02	757	6	1.2	822	2	JT0968	1,4-alpha-glucan bran	2.14e+02
685	6	1.2	649	2	A35816	transcription regulat	2.14e+02	758	6	1.2	835	2	B64689	adenine specific DNA	2.14e+02
686	6	1.2	652	2	B29900	fasciclin I precursor	2.14e+02	759	6	1.2	840	2	A27832	cell division control	2.14e+02
687	6	1.2	657	2	E70529	hypothetical protein	2.14e+02	760	6	1.2	858	2	S30571	DNA topoisomerase (Ar	2.14e+02
688	6	1.2	661	2	B71447	probable GLABRA2 - Ar	2.14e+02	761	6	1.2	859	1	VCLJE1	env polyprotein precu	2.14e+02
689	6	1.2	662	2	E71731	excinuclease ABC chai	2.14e+02	762	6	1.2	859	1	VCLJE2	env polyprotein precu	2.14e+02
690	6	1.2	662	2	E64650	acetyl-CoA synthetase	2.14e+02	763	6	1.2	859	1	VCLJEV	env polyprotein precu	2.14e+02
691	6	1.2	663	2	S20877	dnak-tye molecular c	2.14e+02	764	6	1.2	859	1	VCLJEW	env polyprotein precu	2.14e+02
692	6	1.2	668	2	H71312	probable ATP-dependen	2.14e+02	765	6	1.2	859	1	VCLJE2	env polyprotein precu	2.14e+02
693	6	1.2	668	2	S39836	hypothetical protein	2.14e+02	766	6	1.2	859	1	VCLJE3	env polyprotein precu	2.14e+02
694	6	1.2	669	2	S64795	suppressor protein FP	2.14e+02	767	6	1.2	859	1	VCLJWS	env polyprotein precu	2.14e+02
695	6	1.2	673	2	S59263	probable membrane pro	2.14e+02	768	6	1.2	860	1	VCLJE4	env polyprotein precu	2.14e+02
696	6	1.2	673	2	TVBE15	74.1K protein kinase	2.14e+02	769	6	1.2	862	1	E35905	ATP-dependent Clp pro	2.14e+02
697	6	1.2	674	2	T01309	probable serine/threo	2.14e+02	770	6	1.2	864	2	JH0438	penicillin-binding pr	2.14e+02
698	6	1.2	676	2	WMBEX6	U6 protein - human h	2.14e+02	771	6	1.2	867	2	T00118	hrsh2 protein - sea s	2.14e+02
699	6	1.2	677	2	S33664	flagella-associated p	2.14e+02	772	6	1.2	868	2	S65186	NIP80 protein - yeast	2.14e+02
700	6	1.2	680	4	I38491	nucleophosmin/naplas	2.14e+02	773	6	1.2	878	2	S20486	paramyosin - fruit fl	2.14e+02
701	6	1.2	681	2	A36500	transferrin precursor	2.14e+02	774	6	1.2	879	2	S22028	paramyosin, standard	2.14e+02
702	6	1.2	682	2	S40037	pilJ protein - pseudo	2.14e+02	775	6	1.2	884	2	S53396	telomerase catalytic	2.14e+02
703	6	1.2	683	2	D71680	pilJ protein - E (rne)	2.14e+02	776	6	1.2	885	1	WMBY3L	ribonucleoside-diphos	2.14e+02
704	6	1.2	684	2	E64496	eIF-4A family probabl	2.14e+02	777	6	1.2	892	2	TQ0424	probable transposase	2.14e+02
705	6	1.2	690	2	S75067	H+/K+-exchanging ATPa	2.14e+02	778	6	1.2	902	2	S26002	gene coxi intron 1 pr	2.14e+02
706	6	1.2	691	1	A36295	heat shock transcript	2.14e+02	779	6	1.2	905	2	H71731	DNA gyrase chain A (g	2.14e+02
707	6	1.2	695	2	I54325	gene XE7 protein - hu	2.14e+02	780	6	1.2	917	2	A65000	NADH dehydrogenase (E	2.14e+02
708	6	1.2	697	2	S72353	capsid protein - Sacc	2.14e+02	781	6	1.2	921	2	VGEBEH	glycophorin B precu	2.14e+02
709	6	1.2	700	2	A61527	stonustoxin beta chai	2.14e+02	782	6	1.2	921	2	A33718	retinoblastoma protei	2.14e+02
710	6	1.2	701	2	S61239	hypothetical protein	2.14e+02	783	6	1.2	922	2	S54342	protein-tyrosine-phos	2.14e+02
711	6	1.2	701	2	F70155	Na+/H+ antiporter (na	2.14e+02	784	6	1.2	928	1	RBHU	retinoblastoma-associ	2.14e+02
712	6	1.2	702	1	A48562	coat protein - San Mi	2.14e+02	785	6	1.2	935	2	S57080	hypothetical protein	2.14e+02
713	6	1.2	702	2	A36319	carcinoembryonic anti	2.14e+02	786	6	1.2	935	2	S55051	Bicaudal-C - fruit fl	2.14e+02
714	6	1.2	705	2	S55420	conserved hypothetical	2.14e+02	787	6	1.2	940	2	A40985	Profectin - fruit fly	2.14e+02
715	6	1.2	706	2	T01351	subtilisin-like prote	2.14e+02	788	6	1.2	943	2	T03306	PSD-95/SAP90-associat	2.14e+02
716	6	1.2	706	2	S61717	probable membrane pro	2.14e+02	789	6	1.2	948	2	B71405	retinoblastoma protei	2.14e+02
717	6	1.2	707	2	A42322	ornithine decarboxyla	2.14e+02	790	6	1.2	959	2	A57640	probable kinesin - Ar	2.14e+02
718	6	1.2	723	2	S51788	malate synthase (EC 4	2.14e+02	791	6	1.2	962	2	S03818	carboxymethylcellulas	2.14e+02
719	6	1.2	727	2	S53707	translation initiatio	2.14e+02	792	6	1.2	964	2	S06028	gene suppressor-of-wh	2.14e+02
720	6	1.2	727	2	H69724	DNA topoisomerase III	2.14e+02	793	6	1.2	979	2	A70848	probable membrane pro	2.14e+02
721	6	1.2	729	2	S76065	hypothetical protein	2.14e+02	794	6	1.2	979	2	JC2349	protein-tyrosine-phos	2.14e+02
722	6	1.2	731	2	D71332	probable DNA topoisom	2.14e+02	795	6	1.2	996	2	I48721	PTP 35 protein - mous	2.14e+02
723	6	1.2	734	2	C69691	GTP pyrophosphokinase	2.14e+02	796	6	1.2	1005	2	A64465	hypothetical protein	2.14e+02
724	6	1.2	739	2	I40715	malate synthase (EC 4	2.14e+02	797	6	1.2	1009	2	S64734	retrovirus-related po	2.14e+02
725	6	1.2	739	1	J01893	80.7K alpha trans-ind	2.14e+02	798	6	1.2	1020	1	QEHUH	neurofilament triplet	2.14e+02
726	6	1.2	741	2	F70722	probable glcB protein	2.14e+02	799	6	1.2	1026	2	S51432	hypothetical protein	2.14e+02
727	6	1.2	742	2	S12533	zfa protein - mouse	2.14e+02	800	6	1.2	1028	2	T03516	probable outer membra	2.14e+02
728	6	1.2	742	1	TNBE1	80.7K alpha trans-ind	2.14e+02	801	6	1.2	1030	2	S43211	probable ATP-binding	2.14e+02
729	6	1.2	743	1	DECHE	glutamate dehydrogena	2.14e+02	802	6	1.2	1046	2	S56026	hypothetical protein	2.14e+02
730	6	1.2	743	2	S38143	hypothetical protein	2.14e+02	803	6	1.2	1055	2	JC5216	type I site-specific	2.14e+02
731	6	1.2	746	2	S67203	probable membrane pro	2.14e+02	804	6	1.2	1069	2	A37221	neurofilament triplet	2.14e+02
732	6	1.2	747	2	I48294	gene C1cn4 protein -	2.14e+02	805	6	1.2	1072	2	S77162	DNA topoisomerase (AT	2.14e+02
733	6	1.2	747	1	A57107	kinesin-related prote	2.14e+02	806	6	1.2	1078	2	S77162	119K DNA helicase/pr	2.14e+02
734	6	1.2	748	2	F71927	cag island protein, D	2.14e+02	807	6	1.2	1081	1	WZBEA6	NAD(P)+ transhydrogen	2.14e+02
735	6	1.2	749	2	S50095	translation elongatio	2.14e+02	808	6	1.2	1086	1	G02257	NAD(P)+ transhydrogen	2.14e+02
736	6	1.2	752	2	S40780	CDC28/cdc2-like kinas	2.14e+02	809	6	1.2	1086	2	DEBOXM	NAD(P)+ transhydrogen	2.14e+02
737	6	1.2	754	2	JC5314	protoporphyrin IX mag	2.14e+02	810	6	1.2	1086	2	S54876	NAD(P)+ transhydrogen	2.14e+02
738	6	1.2	758	3	T02925	dipterydyl-peptidase	2.14e+02	811	6	1.2	1087	1	QFM5H	neurofilament triplet	2.14e+02
739	6	1.2	760	2	S23752	iron(II) transport sy	2.14e+02	812	6	1.2	1093	2	I38533	hypothetical protein	2.14e+02
740	6	1.2	773	2	A36932	penicillin amidase (E	2.14e+02	813	6	1.2	1129	2	A29631	phytochrome - oat	2.14e+02
741	6	1.2	774	2	A28392	outer layer protein V	2.14e+02	814	6	1.2	1129	2	S00097	phytochrome 4 - oat	2.14e+02
742	6	1.2	776	2	S03611	outer layer protein V	2.14e+02	815	6	1.2	1129	2	S00096	phytochrome 3 - oat	2.14e+02
743	6	1.2	776	1	VPXRS1	probable tex protein	2.14e+02	816	6	1.2	1129	2	A48843	MHC class II transact	2.14e+02
744	6	1.2	778	2	E71263	probable membrane pro	2.14e+02	817	6	1.2	1130	2	A35098	MHC class III histoco	2.14e+02
745	6	1.2	779	2	S51413	protein kinase DraF-1	2.14e+02	818	6	1.2	1132	2	S60433	probable membrane pro	2.14e+02
746	6	1.2	781	1	TVFFDF	cell surface antigen	2.14e+02	819	6	1.2	1139	2	B54962	sterol regulatory ele	2.14e+02
747	6	1.2	785	2	D71653	probable membrane pro	2.14e+02	820	6	1.2	1141	2	A54962	sterol regulatory ele	2.14e+02
748	6	1.2	790	2	S67803	RNA helicase Gu - hum	2.14e+02	821	6	1.2	1142	2	S50632	protein kinase PAK1 (2.14e+02
749	6	1.2	801	2	PC6010	PML protein, splice f	2.14e+02	822	6	1.2	1148	2	T00016	minor outer capsid pr	2.14e+02
750	6	1.2	802	2	S44382	PML protein, splice f	2.14e+02	823	6	1.2	1164	2	G71827	hypothetical protein	2.14e+02
751	6	1.2	802	2	S42518	U147h protein - Marek	2.14e+02	824	6	1.2	1195	2	C64440	DNA repair protein RA	2.14e+02
752	6	1.2	808	2	JQ2205	plasma (EC 3.4.21.7)	2.14e+02	825	6	1.2	1199	2	G69698	RNA polymerase (beta'	2.14e+02
753	6	1.2	812	1	PLBO			826	6	1.2	1199	2	G69698		

535	6	1.2	463	1	W2BYRS	argininosuccinate lyase	2.14e+02	608	6	1.2	531	2	S76041	hypothetical protein	2.14e+02
536	6	1.2	464	2	S16954	retinoblastoma-bindin	2.14e+02	609	6	1.2	532	2	S33758	flavin-containing mon	2.14e+02
537	6	1.2	465	2	JC1318	triacylglycerol lipase	2.14e+02	610	6	1.2	532	1	A33768	dimethylalanine monoo	2.14e+02
538	6	1.2	466	2	JN0765	alpha-1C-adrenergic r	2.14e+02	611	6	1.2	534	2	S62572	hypothetical protein	2.14e+02
539	6	1.2	466	2	I52862	alpha1C-adrenergic re	2.14e+02	612	6	1.2	534	2	S55635	DNA helicase-primase	2.14e+02
540	6	1.2	466	2	I57959	alpha1C-adrenergic re	2.14e+02	613	6	1.2	535	1	A35182	dimethylalanine monoo	2.14e+02
541	6	1.2	466	2	A35375	alpha-1-adrenergic re	2.14e+02	614	6	1.2	536	1	A35427	dimethylalanine monoo	2.14e+02
542	6	1.2	467	1	PWBSBF	H+-transporting ATP s	2.14e+02	615	6	1.2	539	2	T02705	hypothetical protein	2.14e+02
543	6	1.2	467	1	VHNZ	nucleocapsid protein	2.14e+02	616	6	1.2	542	2	S45557	hypothetical resB hom	2.14e+02
544	6	1.2	467	2	S30839	UTR2 protein - yeast	2.14e+02	617	6	1.2	543	1	KIHUPL	pyruvate kinase (EC 2	2.14e+02
545	6	1.2	467	2	S38458	replication protein A	2.14e+02	618	6	1.2	543	2	A23612	pyruvate kinase (EC 2	2.14e+02
546	6	1.2	468	2	S20951	Na+/H+-exchanging pro	2.14e+02	619	6	1.2	543	1	KIRTP1	pyruvate kinase (EC 2	2.14e+02
547	6	1.2	469	2	JC5741	membrane-bound proton	2.14e+02	620	6	1.2	546	2	S67292	probable membrane pro	2.14e+02
548	6	1.2	469	2	S53810	paracrySTALLINE surfa	2.14e+02	621	6	1.2	547	1	ERADDF4	60.5K fiber protein -	2.14e+02
549	6	1.2	469	2	S17726	H+-transporting ATP s	2.14e+02	622	6	1.2	548	2	I37577	islet cell antigen 51	2.14e+02
550	6	1.2	470	2	A65168	hypothetical 49.9 kD	2.14e+02	623	6	1.2	549	2	S47739	probable alpha, alpha-	2.14e+02
551	6	1.2	470	2	H70795	hypothetical protein	2.14e+02	624	6	1.2	553	2	D69162	pyruvate dehydrogenas	2.14e+02
552	6	1.2	471	2	C70672	hypothetical protein	2.14e+02	625	6	1.2	554	2	JU0188	membrane protein 4.1	2.14e+02
553	6	1.2	471	2	E69435	hypothetical protein	2.14e+02	626	6	1.2	554	2	C64762	probable monooxygenas	2.14e+02
554	6	1.2	472	1	PWBSBM	H+-transporting ATP s	2.14e+02	627	6	1.2	557	2	S62522	hypothetical protein	2.14e+02
555	6	1.2	474	2	S00880	nitrogenase molybdenu	2.14e+02	628	6	1.2	558	1	DEHUE	glutamate dehydrogena	2.14e+02
556	6	1.2	475	2	S20250	splicing factor U2AF	2.14e+02	629	6	1.2	558	1	S03707	glutamate dehydrogena	2.14e+02
557	6	1.2	476	2	JC4646	bone morphogenetic pr	2.14e+02	630	6	1.2	558	1	S16239	glutamate dehydrogena	2.14e+02
558	6	1.2	476	2	C64601	glucosyltransferase -	2.14e+02	631	6	1.2	562	1	ERADN2	60.5K fiber protein -	2.14e+02
559	6	1.2	477	2	S71323	alpha-1A adrenergic r	2.14e+02	632	6	1.2	565	2	T01116	carrot B2 protein hom	2.14e+02
560	6	1.2	478	2	JC4838	bone morphogenetic pr	2.14e+02	633	6	1.2	566	1	HMITV1	hemagglutinin precurs	2.14e+02
561	6	1.2	478	2	A32555	major metazoite surfa	2.14e+02	634	6	1.2	566	1	HMITVD1	hemagglutinin precurs	2.14e+02
562	6	1.2	479	1	VGBEF2	glycoprotein F - huma	2.14e+02	635	6	1.2	566	1	HMITVUR	hemagglutinin precurs	2.14e+02
563	6	1.2	481	2	S69808	lincomycin resistance	2.14e+02	636	6	1.2	566	1	HMITV17	hemagglutinin precurs	2.14e+02
564	6	1.2	482	3	A43828	serine carboxypeptida	2.14e+02	637	6	1.2	566	1	B41648	protein-tyrosine-phos	2.14e+02
565	6	1.2	485	2	JU0300	X-His dipeptidase (EC	2.14e+02	638	6	1.2	566	2	S69888	hemagglutinin precurs	2.14e+02
566	6	1.2	490	2	S70352	protein kinase CLK3 (2.14e+02	639	6	1.2	566	2	S69889	hemagglutinin precurs	2.14e+02
567	6	1.2	490	2	S53639	probable regulatory p	2.14e+02	641	6	1.2	568	2	JC4570	gamma-glutamyltransfe	2.14e+02
568	6	1.2	490	2	E70649	bchx protein - Rhodob	2.14e+02	642	6	1.2	569	1	A05225	gamma-glutamyltransfe	2.14e+02
569	6	1.2	492	2	S32646	regulatory factor U2AF	2.14e+02	643	6	1.2	570	2	S27495	nodu protein - Bradyr	2.14e+02
570	6	1.2	492	2	S32951	phytochrome 5 - oat (2.14e+02	644	6	1.2	573	2	T02405	L-Ornithine N5-oxygen	2.14e+02
571	6	1.2	495	2	S00098	conserved hypothetical	2.14e+02	645	6	1.2	573	2	C71312	probable pyrophosphat	2.14e+02
572	6	1.2	496	2	B64638	spermatogenesis facto	2.14e+02	646	6	1.2	574	1	KIHUPR	pyruvate kinase (EC 2	2.14e+02
573	6	1.2	496	2	A49418	cytochrome-c oxidase	2.14e+02	647	6	1.2	578	2	JC1385	protein kinase (EC 2	2.14e+02
574	6	1.2	499	2	S65657	alpha-1C-adrenergic r	2.14e+02	648	6	1.2	579	2	A55369	tyrosine 3-monooxygen	2.14e+02
575	6	1.2	500	2	A53658	prostaglandin-I synth	2.14e+02	649	6	1.2	581	2	E69322	dolichol-P-glucose sy	2.14e+02
576	6	1.2	500	2	JC2231	sucrose-6-phosphate h	2.14e+02	650	6	1.2	583	2	S50959	probable membrane pro	2.14e+02
577	6	1.2	501	2	S44258	conserved hypothetical	2.14e+02	651	6	1.2	583	2	S19476	hypothetical protein	2.14e+02
578	6	1.2	502	2	E71963	probable cardiolipin	2.14e+02	652	6	1.2	584	2	G70804	hypothetical glycine-	2.14e+02
579	6	1.2	502	2	F64543	glutamate dehydrogena	2.14e+02	653	6	1.2	588	2	A49618	probable ataxia-telan	2.14e+02
580	6	1.2	503	1	DEBOE	hypothetical protein	2.14e+02	654	6	1.2	590	1	A38119	cell division protein	2.14e+02
581	6	1.2	504	2	S23243	hypothetical protein	2.14e+02	655	6	1.2	590	1	A38119	penicillin binding pr	2.14e+02
582	6	1.2	505	2	S67256	probable pyruvate kin	2.14e+02	656	6	1.2	595	2	H71932	hypothetical protein	2.14e+02
583	6	1.2	505	2	T00723	H+-transporting ATP s	2.14e+02	657	6	1.2	599	2	H64579	GTP-binding protein,	2.14e+02
584	6	1.2	508	1	PWZMAM	legumin-like storage	2.14e+02	658	6	1.2	601	2	S12004	tyramine receptor - f	2.14e+02
585	6	1.2	509	2	S18872	transcription initiat	2.14e+02	659	6	1.2	601	2	G71666	nitrogen regulation p	2.14e+02
586	6	1.2	510	2	S41307	hypothetical ABC tran	2.14e+02	660	6	1.2	601	2	JH0150	octopamine receptor t	2.14e+02
587	6	1.2	511	2	S56315	phenylalanine--trNA 1	2.14e+02	661	6	1.2	603	2	H70156	exonuclease ABC, sub	2.14e+02
588	6	1.2	511	2	C69199	histidine kinase sens	2.14e+02	662	6	1.2	603	2	A47545	protein kinase (EC 2	2.14e+02
589	6	1.2	513	2	A70378	serine--trNA ligase (2.14e+02	663	6	1.2	603	2	A54596	protein kinase - mous	2.14e+02
590	6	1.2	513	2	D69016	E2 protein - human pa	2.14e+02	664	6	1.2	603	2	S34130	serine/threonine-spec	2.14e+02
591	6	1.2	514	1	W2WL5	alpha-1B-adrenergic r	2.14e+02	665	6	1.2	603	2	A22282	DNA primase (EC 2.7.7	2.14e+02
592	6	1.2	515	2	JC1525	alpha-1B-adrenergic re	2.14e+02	666	6	1.2	603	2	S76959	GTP-binding membrane	2.14e+02
593	6	1.2	515	2	A40491	env polypeptide - bov	2.14e+02	667	6	1.2	607	2	S42639	ATP-dependent RNA hel	2.14e+02
594	6	1.2	515	2	VCLUB	vacuolar segregation	2.14e+02	668	6	1.2	609	2	A64432	modulation factor pro	2.14e+02
595	6	1.2	515	2	S59811	alpha-1B-adrenergic r	2.14e+02	669	6	1.2	613	2	T00077	gag-like protein - Ch	2.14e+02
596	6	1.2	517	2	JC2332	alpha-1B-adrenergic r	2.14e+02	670	6	1.2	614	2	A25707	type I restriction mo	2.14e+02
597	6	1.2	517	2	A45121	hypothetical protein	2.14e+02	671	6	1.2	616	2	C69226	pyocin SI S1A subunit	2.14e+02
598	6	1.2	521	2	F70772	hypothetical protein	2.14e+02	672	6	1.2	618	2	A36907	72K mitochondrial out	2.14e+02
599	6	1.2	523	2	S47758	hypothetical 59.4K pr	2.14e+02	673	6	1.2	619	2	A36682	hypothetical protein	2.14e+02
600	6	1.2	525	2	JN0443	transcription initiat	2.14e+02	674	6	1.2	622	2	S62532	transpositional regulat	2.14e+02
601	6	1.2	527	1	K1BYD	aspartate kinase (EC	2.14e+02	675	6	1.2	627	2	T02846	hypothetical protein	2.14e+02
602	6	1.2	528	2	JN0445	transcription initiat	2.14e+02	676	6	1.2	630	2	B24584	probable dicarboxylat	2.14e+02
603	6	1.2	528	2	S19366	hypothetical protein	2.14e+02	677	6	1.2	632	2	A71259	probable glucosamine-	2.14e+02
604	6	1.2	529	1	VGNVAC	major envelope glycop	2.14e+02	678	6	1.2	635	2	E71272	dnak-type molecular c	2.14e+02
605	6	1.2	530	2	S12320	PRP9 protein - yeast	2.14e+02	679	6	1.2	637	2	JC5608		2.14e+02
606	6	1.2						680	6	1.2					

389	6	1.2	346	2	E64002	hypothetical protein	2.14e+02	462	6	1.2	401	2	A47258	interleukin-1 beta-co	2.14e+02
390	6	1.2	347	1	C46137	opsin, violet-sensiti	2.14e+02	463	6	1.2	401	2	S63227	hypothetical protein	2.14e+02
391	6	1.2	347	2	S09274	Ig alpha chain C regi	2.14e+02	464	6	1.2	401	2	S77167	isopenicillin N epime	2.14e+02
392	6	1.2	349	2	A34815	carcinoembryonic anti	2.14e+02	465	6	1.2	402	2	JC4382	ribosomal protein L3.	2.14e+02
393	6	1.2	350	2	A69834	conserved hypothetical	2.14e+02	466	6	1.2	402	2	A46495	IL-1 beta convertase	2.14e+02
394	6	1.2	350	2	B38535	A/G-specific adenine	2.14e+02	467	6	1.2	403	1	A53396	flavohe moglobin - A1c	2.14e+02
395	6	1.2	352	2	I50047	rhodopsin - Mexican t	2.14e+02	468	6	1.2	404	2	C70610	probable glucose-1-ph	2.14e+02
396	6	1.2	352	2	S24559	Wnt-2 protein - fruit	2.14e+02	469	6	1.2	406	2	A70015	NADH dehydrogenase ho	2.14e+02
397	6	1.2	353	2	S16785	hemagglutinin - influ	2.14e+02	470	6	1.2	406	2	JC4600	isocitrate dehydrogen	2.14e+02
398	6	1.2	354	2	A35788	hemagglutinin - influ	2.14e+02	471	6	1.2	407	2	JC5632	hemocyanin dioxxygen-b	2.14e+02
399	6	1.2	354	2	E69499	iron-sulfur binding r	2.14e+02	472	6	1.2	407	2	B70733	probable aminotransfe	2.14e+02
400	6	1.2	354	2	I80165	class I histocompatib	2.14e+02	473	6	1.2	407	2	D69316	mRNA 3'-end processin	2.14e+02
401	6	1.2	354	2	A48049	coproporphyrinogen ox	2.14e+02	474	6	1.2	408	2	I48680	probable transposase	2.14e+02
402	6	1.2	354	2	I52444	hypothetical protein	2.14e+02	475	6	1.2	409	2	S60975	hypothetical protein	2.14e+02
403	6	1.2	355	2	S76940	conserved hypothetical	2.14e+02	476	6	1.2	409	2	S60988	hypothetical protein	2.14e+02
404	6	1.2	355	2	F69298	conserved hypothetical	2.14e+02	477	6	1.2	409	2	S60988	serine proteinase inh	2.14e+02
405	6	1.2	355	2	I51319	RH2 opsin - green ano	2.14e+02	478	6	1.2	410	2	I50494	protein kinase MSK-1	2.14e+02
406	6	1.2	355	2	A42347	opsin, green-sensitiv	2.14e+02	479	6	1.2	411	2	S37644	serine proteinase inh	2.14e+02
407	6	1.2	355	2	A46191	iodopsin homolog - to	2.14e+02	480	6	1.2	412	2	S31505	translation elongatio	2.14e+02
408	6	1.2	356	2	G69529	hypothetical protein	2.14e+02	481	6	1.2	413	2	F69007	potassium channel - m	2.14e+02
409	6	1.2	356	2	A47128	carinomycin 4-O-meth	2.14e+02	482	6	1.2	414	2	S48738	transcription factor	2.14e+02
410	6	1.2	358	2	F70583	hypothetical protein	2.14e+02	483	6	1.2	414	2	A53950	transcription factor	2.14e+02
411	6	1.2	358	2	S51263	probable galactosyltr	2.14e+02	484	6	1.2	415	2	I58144	corticotropin-releasi	2.14e+02
412	6	1.2	358	2	S69886	hemagglutinin precurs	2.14e+02	485	6	1.2	415	2	S39535	corticotropin-releasi	2.14e+02
413	6	1.2	360	2	S65210	hypothetical protein	2.14e+02	486	6	1.2	415	2	A56042	mitogen-activated pro	2.14e+02
414	6	1.2	360	2	S74751	CDP-glucose 4,6-dehyd	2.14e+02	487	6	1.2	416	2	A70327	conserved hypothetical	2.14e+02
415	6	1.2	361	2	B46137	opsin, blue-sensitiv	2.14e+02	488	6	1.2	416	2	G02635	ICE-LAP6 - human	2.14e+02
416	6	1.2	362	2	S69197	oleoyl-lacyl-carrier-	2.14e+02	489	6	1.2	417	2	F64660	colicin tolerance-lik	2.14e+02
417	6	1.2	364	1	OOHUG	opsin, green-sensitiv	2.14e+02	490	6	1.2	417	2	G64614	conserved hypothetical	2.14e+02
418	6	1.2	364	1	JT0741	GTP-binding protein 1	2.14e+02	491	6	1.2	417	2	G71854	probable tomb-indepen	2.14e+02
419	6	1.2	364	1	OOHUR	opsin, red-sensitiv	2.14e+02	492	6	1.2	417	2	H71898	hypothetical protein	2.14e+02
420	6	1.2	365	2	S62542	hypothetical protein	2.14e+02	493	6	1.2	417	2	B55473	early switch protein	2.14e+02
421	6	1.2	366	1	DJEC3B	DNA-directed DNA poly	2.14e+02	494	6	1.2	419	2	S68803	probable transcriptio	2.14e+02
422	6	1.2	367	1	A46355	site-specific DNA-met	2.14e+02	495	6	1.2	419	2	G70652	probable sers protein	2.14e+02
423	6	1.2	371	2	D71201	hypothetical protein	2.14e+02	496	6	1.2	419	2	B49418	spermatogenesis facto	2.14e+02
424	6	1.2	372	2	I50492	alpha-1-antitrypsin p	2.14e+02	497	6	1.2	420	2	I51666	Mel-1c receptor subty	2.14e+02
425	6	1.2	372	2	A32375	lymph node homing rec	2.14e+02	498	6	1.2	420	2	E70914	probable lipo protein	2.14e+02
426	6	1.2	375	2	I38879	corticotropin releasi	2.14e+02	499	6	1.2	422	1	A56674	paired box transcript	2.14e+02
427	6	1.2	376	2	B69125	hydrogenase expressio	2.14e+02	500	6	1.2	423	2	A69367	translation elongatio	2.14e+02
428	6	1.2	376	2	S16386	hypothetical protein	2.14e+02	501	6	1.2	423	2	F64436	hypothetical protein	2.14e+02
429	6	1.2	376	2	S46032	SUR1 protein homolog	2.14e+02	502	6	1.2	424	2	B69210	inward rectifier pota	2.14e+02
430	6	1.2	377	2	A57511	interleukin-1 beta co	2.14e+02	503	6	1.2	424	2	I38979	conserved hypothetical	2.14e+02
431	6	1.2	377	2	S74628	hypothetical protein	2.14e+02	504	6	1.2	425	2	S52852	inward rectifier pota	2.14e+02
432	6	1.2	378	2	S12190	mobl protein - Thiba	2.14e+02	505	6	1.2	425	2	C64567	fucoyltransferase -	2.14e+02
433	6	1.2	379	2	B69344	hypothetical protein	2.14e+02	506	6	1.2	425	2	I48202	potassium channel-lik	2.14e+02
434	6	1.2	380	2	A70646	probable PPE protein	2.14e+02	507	6	1.2	426	2	D64970	hypothetical protein	2.14e+02
435	6	1.2	381	2	C64416	conserved hypothetical	2.14e+02	508	6	1.2	427	2	S19338	hypothetical protein	2.14e+02
436	6	1.2	382	2	E36607	nucleocapsid protein	2.14e+02	509	6	1.2	429	2	B65656	alpha-1C-adrenergic r	2.14e+02
437	6	1.2	382	2	S03762	nucleocapsid protein	2.14e+02	510	6	1.2	433	2	B69495	aconitase (acn) homol	2.14e+02
438	6	1.2	382	2	S36476	E2 protein - human pa	2.14e+02	511	6	1.2	433	2	B69115	coenzyme F390 synthet	2.14e+02
439	6	1.2	382	1	S24282	nucleocapsid protein	2.14e+02	512	6	1.2	435	2	S19493	hypothetical protein	2.14e+02
440	6	1.2	382	1	S33573	ubiquinol--cytochrome	2.14e+02	513	6	1.2	436	2	G71862	alpha-(1,3)-fucoyltr	2.14e+02
441	6	1.2	382	2	S03936	nucleocapsid protein	2.14e+02	514	6	1.2	436	2	S42234	paired box transcript	2.14e+02
442	6	1.2	382	2	S47428	nucleocapsid protein	2.14e+02	515	6	1.2	440	1	B71293	hypothetical protein	2.14e+02
443	6	1.2	382	2	VHIHPC	nucleocapsid protein	2.14e+02	516	6	1.2	441	2	JQ2191	nucleocapsid protein	2.14e+02
444	6	1.2	382	1	KIECGG	galactokinase (EC 2.7	2.14e+02	517	6	1.2	442	2	H71410	probable RNA helicase	2.14e+02
445	6	1.2	385	2	I84552	prophage Sf6-like int	2.14e+02	518	6	1.2	444	2	S11712	transcription initiat	2.14e+02
446	6	1.2	386	2	H71701	succinyl-CoA syntheta	2.14e+02	519	6	1.2	444	2	A48260	corticoliberin recept	2.14e+02
447	6	1.2	388	2	H65126	probable general secr	2.14e+02	520	6	1.2	445	2	T01591	hypothetical protein	2.14e+02
448	6	1.2	388	2	A55597	oxytocin receptor - r	2.14e+02	521	6	1.2	448	2	A71347	hypothetical protein	2.14e+02
449	6	1.2	393	2	A55859	regulatory protein na	2.14e+02	522	6	1.2	448	1	A56018	triacylglycerol lipas	2.14e+02
450	6	1.2	393	2	E64485	phosphoribosylglycina	2.14e+02	523	6	1.2	449	1	S30205	transcription factor	2.14e+02
451	6	1.2	393	2	G70891	ketol-acid reductoiso	2.14e+02	524	6	1.2	451	1	A40168	protein-tyrosine kina	2.14e+02
452	6	1.2	395	2	A24709	modulation protein no	2.14e+02	525	6	1.2	451	2	S49016	hypothetical protein	2.14e+02
453	6	1.2	395	1	JQ0396	aurora-related kinase	2.14e+02	526	6	1.2	451	2	S58653	coproporphyrinogen ox	2.14e+02
454	6	1.2	395	2	JC5975	hypothetical protein	2.14e+02	527	6	1.2	454	2	I37257	alpha (1,3)-fucoyltr	2.14e+02
455	6	1.2	396	2	QOECTR	tyrosine transaminase	2.14e+02	528	6	1.2	454	2	B71914	signal recognition pa	2.14e+02
456	6	1.2	396	2	S76602	hypothetical protein	2.14e+02	529	6	1.2	454	2	E70448	hypothetical protein	2.14e+02
457	6	1.2	397	1	XNECY	interleukin-1 beta co	2.14e+02	530	6	1.2	460	2	H70924	sulfite oxidase (EC 1	2.14e+02
458	6	1.2	397	2	S76786	coat protein precurs	2.14e+02	531	6	1.2	460	2	A34180	E2 protein - human pa	2.14e+02
459	6	1.2	399	1	VCBAND	bombesin-like peptide	2.14e+02	532	6	1.2	461	2	S36593	argininosuccinate lya	2.14e+02
460	6	1.2	399	1	A46632			533	6	1.2	463	2	S43539		
461	6	1.2	399	2				534	6	1.2	463	2			

243	6	1.2	210	2	A71876	probable aldehyde deh	2.14e+02	316	6	1.2	288	1	P3VXPS	3a protein - peanut s	2.14e+02
244	6	1.2	212	2	I67437	cysteine proteinase (2.14e+02	317	6	1.2	289	2	S56153	H+-transporting ATP s	2.14e+02
245	6	1.2	214	4	S58320	hypothetical protein	2.14e+02	318	6	1.2	289	2	F64374	modification methylas	2.14e+02
246	6	1.2	216	2	I38470	olfactory receptor -	2.14e+02	319	6	1.2	289	2	S56287	hypothetical protein	2.14e+02
247	6	1.2	216	2	I38474	olfactory receptor -	2.14e+02	320	6	1.2	290	2	S47839	hypothetical protein	2.14e+02
248	6	1.2	216	2	S37349	luxM protein - Vibrio	2.14e+02	321	6	1.2	290	2	H70057	agmatinase homolog yw	2.14e+02
249	6	1.2	216	2	S61544	transferrin binding p	2.14e+02	322	6	1.2	290	2	S39569	syntaxin-related prot	2.14e+02
250	6	1.2	216	2	D70813	probable NarL - Mycob	2.14e+02	323	6	1.2	293	2	I40182	hydroxyquinol 1,2-dio	2.14e+02
251	6	1.2	216	2	S08437	vif protein - human i	2.14e+02	324	6	1.2	294	2	T01307	splicing factor SF-2	2.14e+02
252	6	1.2	219	2	S59134	probable transcriptio	2.14e+02	325	6	1.2	294	2	JT0536	ADP-ribosyl-nitrogena	2.14e+02
253	6	1.2	223	2	E69125	DNA-dependent DNA pol	2.14e+02	326	6	1.2	296	2	A45447	zinc finger (alternat	2.14e+02
254	6	1.2	224	2	I78853	reI2 - rat (fragment)	2.14e+02	327	6	1.2	296	1	OPNE7	peroxidase (EC 1.11.1	2.14e+02
255	6	1.2	225	2	S23568	adenylate kinase (EC	2.14e+02	328	6	1.2	297	2	J01205	attachment protein -	2.14e+02
256	6	1.2	227	2	B35989	nodulation protein W	2.14e+02	329	6	1.2	298	1	MGNZ	major surface glycopr	2.14e+02
257	6	1.2	229	2	S77621	late competence gene	2.14e+02	330	6	1.2	298	2	A71159	probable membrane pro	2.14e+02
258	6	1.2	229	2	B69712	mutants block sporula	2.14e+02	331	6	1.2	298	1	MGNZRL	major surface glycopr	2.14e+02
259	6	1.2	231	2	E71678	hypothetical protein	2.14e+02	332	6	1.2	298	2	G64390	coenzyme F420 hydroge	2.14e+02
260	6	1.2	232	2	B71536	probable DNA pol III	2.14e+02	333	6	1.2	299	2	G71088	hypothetical protein	2.14e+02
261	6	1.2	233	2	D70058	hypothetical protein	2.14e+02	334	6	1.2	301	2	T03320	major capsid protein	2.14e+02
262	6	1.2	233	1	PQBPL	replication protein P	2.14e+02	335	6	1.2	301	1	VHBPL	major capsid protein	2.14e+02
263	6	1.2	234	2	S41906	COR1 protein - golden	2.14e+02	336	6	1.2	303	2	F70356	transcription regulat	2.14e+02
264	6	1.2	237	2	B71424	hypothetical protein	2.14e+02	337	6	1.2	303	2	S74949	hypothetical protein	2.14e+02
265	6	1.2	238	2	B64074	purine-nucleoside pho	2.14e+02	338	6	1.2	304	2	S67148	hypothetical protein	2.14e+02
266	6	1.2	243	2	S56405	hypothetical 26.6k pr	2.14e+02	339	6	1.2	306	2	B71439	hypothetical protein	2.14e+02
267	6	1.2	243	2	JE0204	testicular protein Tp	2.14e+02	340	6	1.2	306	2	S17713	transcription regulat	2.14e+02
268	6	1.2	244	2	JC2379	cell-specific helix-1	2.14e+02	341	6	1.2	309	2	E69347	conserved hypothetical	2.14e+02
269	6	1.2	244	2	S73539	triophosphate isome	2.14e+02	342	6	1.2	309	2	A43865	conserved hypothetical	2.14e+02
270	6	1.2	246	2	I51325	proteolipid protein D	2.14e+02	343	6	1.2	310	2	T01224	probable Cl-tetrahydr	2.14e+02
271	6	1.2	247	2	A69376	hypothetical protein	2.14e+02	344	6	1.2	311	2	H64157	suffi protein homolog	2.14e+02
272	6	1.2	248	2	E70600	hypothetical protein	2.14e+02	345	6	1.2	311	2	H69860	conserved hypothetical	2.14e+02
273	6	1.2	248	1	JH0612	amphiregulin precursor	2.14e+02	346	6	1.2	312	2	A40708	basic-helix-loop-heli	2.14e+02
274	6	1.2	249	3	T00169	anti repressor - Strap	2.14e+02	347	6	1.2	312	2	JE0328	hypothetical protein	2.14e+02
275	6	1.2	249	2	S76255	hypothetical protein	2.14e+02	348	6	1.2	313	2	T01587	olfactory receptor -	2.14e+02
276	6	1.2	251	2	D69919	phage-related DNA-bin	2.14e+02	349	6	1.2	313	2	S20572	nickel transport syst	2.14e+02
277	6	1.2	252	2	S76336	conserved hypothetical	2.14e+02	350	6	1.2	314	2	S47696	coat protein VP2 - mo	2.14e+02
278	6	1.2	253	2	BVECC2	hypothetical protein	2.14e+02	351	6	1.2	314	2	S47696	coat protein VP2 - mo	2.14e+02
279	6	1.2	253	1	I38427	probable sulfate tran	2.14e+02	352	6	1.2	319	1	VVVP2	transcription initiat	2.14e+02
280	6	1.2	254	2	I38427	4-1BB ligand - human	2.14e+02	353	6	1.2	319	1	C44020	TrbB - plasmid RK2	2.14e+02
281	6	1.2	256	2	A32373	transformer-2 protein	2.14e+02	354	6	1.2	320	2	S69547	transcription initiat	2.14e+02
282	6	1.2	257	2	S78200	H+-transporting ATP s	2.14e+02	355	6	1.2	320	2	C71071	hypothetical protein	2.14e+02
283	6	1.2	257	2	A55925	synaptonemal complex	2.14e+02	356	6	1.2	322	2	G64949	probable S-adenosylme	2.14e+02
284	6	1.2	260	2	D27793	methyl coenzyme M red	2.14e+02	357	6	1.2	324	2	S39502	vegetative storage pr	2.14e+02
285	6	1.2	260	2	E69956	phosphate ABC transpo	2.14e+02	358	6	1.2	324	2	G69515	transcription regulat	2.14e+02
286	6	1.2	261	2	S74600	hypothetical protein	2.14e+02	359	6	1.2	325	2	JC5276	HXC-26 protein - huma	2.14e+02
287	6	1.2	261	2	D64075	ureH protein homolog	2.14e+02	360	6	1.2	326	2	JQ1437	hemagglutinin - influ	2.14e+02
288	6	1.2	262	2	G69018	arylsulfatase regulat	2.14e+02	361	6	1.2	327	2	JC4586	G-protein-gated inwar	2.14e+02
289	6	1.2	263	2	A69066	cation transporting P	2.14e+02	362	6	1.2	327	2	S33670	chain length determin	2.14e+02
290	6	1.2	263	2	C64944	transcription regulat	2.14e+02	363	6	1.2	328	2	I39492	nifR3 protein - Azosp	2.14e+02
291	6	1.2	263	2	A35147	hypothetical protein	2.14e+02	364	6	1.2	332	2	A69224	tungsten formylmethan	2.14e+02
292	6	1.2	264	2	A31638	transformer-2 sex-det	2.14e+02	365	6	1.2	334	2	S12742	transcription factor	2.14e+02
293	6	1.2	265	2	S70247	hypothetical protein	2.14e+02	366	6	1.2	334	2	A71169	probable flagellin B	2.14e+02
294	6	1.2	266	2	S50189	phosphotransferase sy	2.14e+02	367	6	1.2	334	2	S25785	hypothetical protein	2.14e+02
295	6	1.2	268	2	G71271	probable ABC transpor	2.14e+02	368	6	1.2	334	2	S25785	probable gpda2 protei	2.14e+02
296	6	1.2	268	2	C37280	C/EBP-related protein	2.14e+02	369	6	1.2	334	2	C70673	transcription factor	2.14e+02
297	6	1.2	268	2	B37279	enhancer-binding prot	2.14e+02	370	6	1.2	336	2	TVMSJA	probable membrane pro	2.14e+02
298	6	1.2	269	1	OTNC3	cytochrome-c oxidase	2.14e+02	371	6	1.2	336	2	S55863	conserved hypothetical	2.14e+02
299	6	1.2	269	2	C69312	molybdopterin oxidore	2.14e+02	372	6	1.2	338	2	S55657	capsid protein 62 - e	2.14e+02
300	6	1.2	269	2	S65034	cytochrome-c oxidase	2.14e+02	373	6	1.2	338	2	S37920	MBR1 protein precursor	2.14e+02
301	6	1.2	272	2	S58290	invasion-associated p	2.14e+02	374	6	1.2	340	2	PSECL2	lysophospholipase (EC	2.14e+02
302	6	1.2	276	2	I39651	probable transposase	2.14e+02	375	6	1.2	341	2	I55623	thromboxane A2 recept	2.14e+02
303	6	1.2	277	2	I39005	cysteine proteinase (2.14e+02	376	6	1.2	342	2	A70152	hypothetical protein	2.14e+02
304	6	1.2	277	2	A55315	cysteine proteinase (2.14e+02	377	6	1.2	342	2	S29894	strictosidine synthas	2.14e+02
305	6	1.2	277	2	JC5410	CP32 protein - mouse	2.14e+02	378	6	1.2	343	2	A41874	transcription repress	2.14e+02
306	6	1.2	277	2	S64710	cysteine proteinase (2.14e+02	379	6	1.2	343	2	JQ2371	hemagglutinin - influ	2.14e+02
307	6	1.2	277	2	A31840	RNA polymerase beta c	2.14e+02	380	6	1.2	343	2	JQ2372	hemagglutinin - influ	2.14e+02
308	6	1.2	278	2	D70814	probable fadD16 prote	2.14e+02	381	6	1.2	343	2	JQ2370	hemagglutinin - influ	2.14e+02
309	6	1.2	280	2	T03543	probable cobalt trans	2.14e+02	382	6	1.2	343	1	HMIWTA	hemagglutinin HAI - i	2.14e+02
310	6	1.2	281	2	H71308	hypothetical protein	2.14e+02	383	6	1.2	344	2	S01325	hemagglutinin - influ	2.14e+02
311	6	1.2	281	2	S08410	probable transcriptio	2.14e+02	384	6	1.2	344	2	P00321	hemagglutinin - influ	2.14e+02
312	6	1.2	283	2	D64448	hypothetical protein	2.14e+02	385	6	1.2	344	2	P00319	hemagglutinin - influ	2.14e+02
313	6	1.2	284	2	A35419	neutrophil protein -	2.14e+02	386	6	1.2	344	2	S35140	probable ketol-acid r	2.14e+02
314	6	1.2	285	2	G31482	H+-transporting ATP s	2.14e+02	387	6	1.2	344	2	JQ1643	hemagglutinin - influ	2.14e+02
315	6	1.2	286	1	MNIVC7	nonstructural protein	2.14e+02	388	6	1.2	346	2	I40476	conserved hypothetical	2.14e+02

97	7	1.4	502	2	S36494	E2 protein - human pa	4.44e+00	170	6	1.2	114	2	SL9433	hypothetical protein	2.14e+02
98	7	1.4	514	1	W2WLB5	E2 protein - human pa	4.44e+00	171	6	1.2	117	1	WMBP5B	gene s protein - phag	2.14e+02
99	7	1.4	516	2	S44191	serine-type carboxype	4.44e+00	172	6	1.2	117	2	D69335	succinate dehydrogena	2.14e+02
100	7	1.4	524	2	I55595	splicing factor - hum	4.44e+00	173	6	1.2	118	2	S27476	hypothetical protein	2.14e+02
101	7	1.4	555	2	S69641	hypothetical protein	4.44e+00	174	6	1.2	118	2	S75220	hypothetical protein	2.14e+02
102	7	1.4	615	2	A53616	lamin B receptor - hu	4.44e+00	175	6	1.2	121	1	R5EC7	ribosomal protein L7/	2.14e+02
103	7	1.4	620	2	JC5567	lamin B receptor - ra	4.44e+00	176	6	1.2	121	1	R5EB12	ribosomal protein L7/	2.14e+02
104	7	1.4	620	2	F64408	coenzyme F420 hydroge	4.44e+00	177	6	1.2	125	1	F1SP4	photosystem I chain I	2.14e+02
105	7	1.4	634	2	A64521	outer membrane protei	4.44e+00	178	6	1.2	126	2	SL6481	hypothetical protein	2.14e+02
106	7	1.4	637	2	A36427	lamin B receptor - ch	4.44e+00	179	6	1.2	131	2	PC4345	phosphoinositide 3-ki	2.14e+02
107	7	1.4	646	2	S47929	signal recognition pa	4.44e+00	180	6	1.2	131	1	BGB02	spermatid transiltion	2.14e+02
108	7	1.4	656	2	S76505	hypothetical protein	4.44e+00	181	6	1.2	131	2	S33375	homeotic protein Abd-	2.14e+02
109	7	1.4	668	2	A71986	probable outer membra	4.44e+00	182	6	1.2	133	2	B69162	conserved hypothetical	2.14e+02
110	7	1.4	688	2	S61249	hypothetical protein	4.44e+00	183	6	1.2	133	2	SI0038	hypothetical protein	2.14e+02
111	7	1.4	692	2	S37976	interferon-induced vi	4.44e+00	184	6	1.2	135	2	I64147	lactoylglytathione ly	2.14e+02
112	7	1.4	715	2	B33481	hypothetical protein	4.44e+00	185	6	1.2	135	2	S69464	hypothetical protein	2.14e+02
113	7	1.4	759	2	S50544	hypothetical protein	4.44e+00	186	6	1.2	136	2	I53300	interleukin-1-beta-co	2.14e+02
114	7	1.4	769	2	I56546	Shaw type potassium c	4.44e+00	187	6	1.2	136	2	D35846	transformer-2 sex-det	2.14e+02
115	7	1.4	814	2	I40048	S-layer protein - precu	4.44e+00	188	6	1.2	138	1	BGHU2	spermatid transition	2.14e+02
116	7	1.4	901	2	S50987	SOKI protein - yeast	4.44e+00	189	6	1.2	139	2	A38612	insulin-like growth f	2.14e+02
117	7	1.4	1032	2	A57514	RNA helicase HEL117 -	4.44e+00	190	6	1.2	139	2	D70756	hypothetical protein	2.14e+02
118	7	1.4	1044	2	T02615	probable glycine dehy	4.44e+00	191	6	1.2	141	2	D71530	probable L11 ribosoma	2.14e+02
119	7	1.4	1125	2	S67794	probable membrane pro	4.44e+00	192	6	1.2	144	2	A64662	ATP synthase F0, subu	2.14e+02
120	7	1.4	1150	2	S49956	probable membrane pro	4.44e+00	193	6	1.2	145	2	JC2575	hypothetical protein	2.14e+02
121	7	1.4	1203	2	S26650	DNA-binding protein 5	4.44e+00	194	6	1.2	146	2	S53702	hypothetical protein	2.14e+02
122	7	1.4	1220	2	A56236	probable RNA helicase	4.44e+00	195	6	1.2	148	2	F71367	probable PTS system,	2.14e+02
123	7	1.4	1244	2	S25327	cytoskeleton assembly	4.44e+00	196	6	1.2	152	2	S53640	hypothetical protein	2.14e+02
124	7	1.4	1246	2	S60954	probable membrane pro	4.44e+00	197	6	1.2	152	2	C71973	protein kinase c1k3-1	2.14e+02
125	7	1.4	1257	2	T00486	probable serine/threo	4.44e+00	198	6	1.2	153	2	JC5854	polyketide synthase (2.14e+02
126	7	1.4	1280	2	T00365	hypothetical protein	4.44e+00	199	6	1.2	155	2	A64533	conserved hypothetical	2.14e+02
127	7	1.4	1339	2	JC4387	epidermal growth fact	4.44e+00	200	6	1.2	155	2	A49547	nucleoside-diphosphat	2.14e+02
128	7	1.4	1342	2	A36223	kinase-related transf	4.44e+00	201	6	1.2	156	2	B45447	zinc finger (alternat	2.14e+02
129	7	1.4	1342	2	I59164	HER3 protein precursor	4.44e+00	202	6	1.2	157	2	C70080	conserved hypothetical	2.14e+02
130	7	1.4	1507	2	B47328	natural killer cell t	4.44e+00	203	6	1.2	158	2	S28744	NADH dehydrogenase (u	2.14e+02
131	7	1.4	1791	2	T02345	hypothetical protein	4.44e+00	204	6	1.2	161	2	S01149	H+-transporting ATP s	2.14e+02
132	7	1.4	1926	2	JC4842	DNA-binding nuclear p	4.44e+00	205	6	1.2	161	2	C70613	probable seCF - Mycob	2.14e+02
133	7	1.4	2161	2	JH0564	calcium channel alpha	4.44e+00	206	6	1.2	163	2	I38191	nucleic acid binding	2.14e+02
134	7	1.4	2181	2	A38198	calcium channel alpha	4.44e+00	207	6	1.2	163	2	S61520	hemoglobin protein A	2.14e+02
135	7	1.4	2278	2	S56274	FAB1 protein - yeast	4.44e+00	208	6	1.2	163	2	SL4572	hemoglobin protein A	2.14e+02
136	7	1.4	4128	2	JC6306	protein kinase (EC 2.	4.44e+00	209	6	1.2	170	2	S05303	plastocyanin precursor	2.14e+02
137	6	1.2	14	2	B56884	Pax-QNR, long form -	2.14e+02	210	6	1.2	173	2	E69183	hypothetical protein	2.14e+02
138	6	1.2	20	2	S28405	lamin B receptor - tu	2.14e+02	211	6	1.2	175	2	S36749	transcription factor	2.14e+02
139	6	1.2	31	2	A41269	protamine Z3 - smalle	2.14e+02	212	6	1.2	177	2	S39519	H+-transporting ATP s	2.14e+02
140	6	1.2	35	2	B45316	sperm-specific protam	2.14e+02	213	6	1.2	178	2	B69944	hypothetical protein	2.14e+02
141	6	1.2	36	2	A37172	collagen alpha 1(XII)	2.14e+02	214	6	1.2	178	2	S26044	sex-determining prote	2.14e+02
142	6	1.2	37	2	S29829	protamine Z3 - smalle	2.14e+02	215	6	1.2	180	2	H64333	hypothetical protein	2.14e+02
143	6	1.2	40	2	B27740	gas-vesicle protein -	2.14e+02	216	6	1.2	180	2	S56056	ribosomal protein L18	2.14e+02
144	6	1.2	41	2	I68638	retinoblastoma suscep	2.14e+02	217	6	1.2	180	2	T01045	hypothetical protein	2.14e+02
145	6	1.2	46	2	G02087	RNA binding motif pro	2.14e+02	218	6	1.2	181	2	A40607	monofunctional choris	2.14e+02
146	6	1.2	55	1	VHNVAC	nucleocapsid protein	2.14e+02	219	6	1.2	183	2	E70588	hypothetical protein	2.14e+02
147	6	1.2	65	1	VHNVBM	nucleocapsid protein	2.14e+02	220	6	1.2	184	2	S46808	hypothetical protein	2.14e+02
148	6	1.2	69	2	S39424	protamine p1 - Austr	2.14e+02	221	6	1.2	184	2	S26046	sex-determining prote	2.14e+02
149	6	1.2	76	2	S28127	gas-vesicle protein g	2.14e+02	222	6	1.2	185	2	F71349	probable transcriptio	2.14e+02
150	6	1.2	78	2	A70946	gas-vesicle protein -	2.14e+02	223	6	1.2	186	2	A70109	hypothetical protein	2.14e+02
151	6	1.2	83	2	G64756	probable ribosomal pr	2.14e+02	224	6	1.2	186	2	S74835	hypothetical protein	2.14e+02
152	6	1.2	84	2	D28928	ykxi protein precursor	2.14e+02	225	6	1.2	187	1	QOEC30	prepilin peptidase de	2.14e+02
153	6	1.2	85	2	S07323	pregnancy-specific be	2.14e+02	226	6	1.2	188	1	R5RT81	ribosomal protein L18	2.14e+02
154	6	1.2	85	2	S39512	DNA-directed RNA poly	2.14e+02	227	6	1.2	188	2	S38352	ribosomal protein L18	2.14e+02
155	6	1.2	87	2	A55571	gas-vesicle protein -	2.14e+02	228	6	1.2	189	2	I65744	gene MAC30 protein -	2.14e+02
156	6	1.2	91	2	S59647	chloride conductance	2.14e+02	229	6	1.2	192	2	E71081	hypothetical protein	2.14e+02
157	6	1.2	92	2	S34115	hypothetical protein	2.14e+02	230	6	1.2	195	2	A64377	fumarate hydratase (E	2.14e+02
158	6	1.2	95	2	S56644	sperm-specific protei	2.14e+02	231	6	1.2	195	2	B36970	mannitol operon repre	2.14e+02
159	6	1.2	99	2	A56644	inverted repeat compo	2.14e+02	232	6	1.2	196	2	D64918	glucuronide repressor	2.14e+02
160	6	1.2	99	1	CUSU	plastocyanin - shephe	2.14e+02	233	6	1.2	197	2	A29648	female-specific trans	2.14e+02
161	6	1.2	99	1	CUPU	trypsin inhibitor 2 p	2.14e+02	234	6	1.2	199	2	S41316	coat protein - cucurb	2.14e+02
162	6	1.2	99	2	S65661	trypsin inhibitor 2 p	2.14e+02	235	6	1.2	199	2	JU0052	puromycin N-acetyltra	2.14e+02
163	6	1.2	100	2	S50733	hypothetical protein	2.14e+02	236	6	1.2	200	1	R3MD4	ribosomal protein S4,	2.14e+02
164	6	1.2	101	2	D71426	hypothetical protein	2.14e+02	237	6	1.2	201	2	S60949	proteinase inhibitor	2.14e+02
165	6	1.2	101	2	S51384	hypothetical protein	2.14e+02	238	6	1.2	203	2	S60449	probable membrane pro	2.14e+02
166	6	1.2	101	2	D69395	H+-transporting ATP s	2.14e+02	239	6	1.2	204	2	T02386	hypothetical protein	2.14e+02
167	6	1.2	103	2	D70159	ribosomal protein S10	2.14e+02	240	6	1.2	205	2	SL9634	nucleolar protein GAR	2.14e+02
168	6	1.2	104	1	BVECAA	arta protein - Escher	2.14e+02	241	6	1.2	209	2	S35108	cell division protein	2.14e+02
169	6	1.2	106	2	S32964	hypothetical protein	2.14e+02	242	6	1.2	209	1	S48459	probable dual specifi	2.14e+02



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSPrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Aug 10 11:42:14 1999; MasPar time 21.14 Seconds
953.603 Million cell updates/sec

Tabular output not generated.

Title: >US-08-287-669-19
Description: (1-503) from US08287669.pep
Perfect Score: 503
Sequence: 1 MMRQDRSLRLRNIMFSSH.....MTSRLKKFYFWPEARNSAV 503

Scoring table: TABLE unitprotatable
Gap 60

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 1000 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 3.750; Variance 0.445; scale 8.419

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	452	89.9	503	2	A49429	interleukin-1 beta-co
2	8	1.6	198	2	I48723	ciliary neurotrophic
3	8	1.6	199	1	UNRBCF	ciliary neurotrophic
4	8	1.6	200	1	UNRTCF	ciliary neurotrophic
5	8	1.6	264	2	JC6125	U2 small nuclear ribo
6	8	1.6	301	2	A70039	ABC transporter (ATP-
7	8	1.6	495	2	S75895	hypothetical protein
8	8	1.6	506	1	W2WL47	E2 protein - human pa
9	8	1.6	821	2	S39983	eps8 protein - mouse
10	8	1.6	822	2	I38728	epidermal growth fact
11	8	1.4	34	2	S21080	sperm protein EMI - m
12	8	1.4	56	2	C58213	protamine II - Americ
13	8	1.4	58	2	S34045	protamine I - North Ame
14	8	1.4	58	2	A58208	protamine I-1 - paint
15	8	1.4	61	2	S39425	protamine P1 - duckbl
16	8	1.4	62	2	B58213	protamine I - America
17	8	1.4	101	2	S65494	sperm-specific protei
18	8	1.4	135	2	A36828	orf10 protein - Autogr
19	8	1.4	147	2	I51168	hemoglobin beta-4B ch
20	8	1.4	148	2	A45316	sperm-specific histon
21	8	1.4	158	3	B45317	PL-II* and PL-IV comm
22	8	1.4	162	1	Q0BE19	BLRF2 protein - human
23	8	1.4	164	2	S14016	X16 protein - mouse

24	7	1.4	164	2	I54089	pre-mRNA splicing fac
25	7	1.4	175	2	A39171	T-cell surface glycop
26	7	1.4	198	2	S53911	ribosomal protein L16
27	7	1.4	199	2	S26047	transformer sex-deter
28	7	1.4	201	2	S26045	sex-determining prote
29	7	1.4	202	2	A42410	ryudocan precursor -
30	7	1.4	203	2	S63633	hypothetical protein
31	7	1.4	203	3	A45317	PL-II* and PL-IV comm
32	7	1.4	204	3	S35256	actin (clone Tac9) -
33	7	1.4	204	2	F36950	urease accessory prot
34	7	1.4	218	2	B70645	probable fuculose-1-
35	7	1.4	221	2	B42701	PR264 protein - chick
36	7	1.4	221	2	A42634	splicing factor SC35
37	7	1.4	221	2	A42701	PR264/SC35 protein -
38	7	1.4	238	2	A57198	splicing factor, argi
39	7	1.4	240	2	A46179	U2 snRNP auxiliary fa
40	7	1.4	244	2	S44822	F44E2.3 protein - Cae
41	7	1.4	248	2	A40040	alternative splicing
42	7	1.4	252	2	C43679	ORF3 protein - Autogr
43	7	1.4	253	2	E59098	phosphate transport s
44	7	1.4	268	2	S61003	hypothetical protein
45	7	1.4	269	2	B47112	growth response prote
46	7	1.4	272	2	S59042	splicing factor SRP40
47	7	1.4	274	2	A55335	myelin regulatory fac
48	7	1.4	288	2	S68798	RNA-binding protein S
49	7	1.4	299	1	AHRB	Ig alpha chain C regi
50	7	1.4	303	2	F71185	splicing factor SF-2
51	7	1.4	310	2	F64146	hypothetical protein
52	7	1.4	315	2	C64982	hypothetical 35.2 kD
53	7	1.4	315	1	MNKR4S	nonstructural protein
54	7	1.4	334	1	W2WLB4	E2 protein - bovine p
55	7	1.4	335	2	S09275	Ig alpha chain C regi
56	7	1.4	338	2	S09276	Ig alpha chain C regi
57	7	1.4	339	2	S09264	Ig alpha chain C regi
58	7	1.4	343	2	S09272	Ig alpha chain C regi
59	7	1.4	344	2	S59043	spilling factor SRP55
60	7	1.4	346	2	S48380	IMP2 protein - yeast
61	7	1.4	348	2	S09273	Ig alpha chain C regi
62	7	1.4	348	2	S09270	Ig alpha chain C regi
63	7	1.4	350	2	A40459	nuclear phosphoprotei
64	7	1.4	352	2	S09266	Ig alpha chain C regi
65	7	1.4	357	2	S09269	Ig alpha chain C regi
66	7	1.4	357	2	S09265	Ig alpha chain C regi
67	7	1.4	357	2	S09267	Ig alpha chain C regi
68	7	1.4	358	2	S09268	Ig alpha chain C regi
69	7	1.4	360	2	S09271	Ig alpha chain C regi
70	7	1.4	364	2	S62183	cysteine proteinase T
71	7	1.4	373	2	S47911	PDS1 protein - yeast
72	7	1.4	376	2	A37282	52K active chromatin
73	7	1.4	376	1	A48681	placental thrombin in
74	7	1.4	378	2	A57488	proteinase inhibitor
75	7	1.4	379	2	S49999	40K protein - frog vi
76	7	1.4	379	2	S55421	acyl-CoA dehydrogenas
77	7	1.4	396	2	S58223	LSRI protein - yeast
78	7	1.4	410	2	S18157	globulin 2 - eastern
79	7	1.4	418	2	B57511	interleukin-1 beta co
80	7	1.4	429	2	D70468	conserved hypothetical
81	7	1.4	431	2	I76773	ARP binding component
82	7	1.4	438	2	B69023	conserved hypothetical
83	7	1.4	440	2	B70326	conserved hypothetical
84	7	1.4	448	2	A36311	70K U1 small nuclear
85	7	1.4	452	2	S36482	E2 protein - human pa
86	7	1.4	456	2	S50158	hypothetical protein
87	7	1.4	477	2	S22027	paramyosin, mini - fir
88	7	1.4	478	1	UFPSDF	aspartate ammonia-lya
89	7	1.4	483	2	PNO099	son3 protein - human
90	7	1.4	483	2	S36470	E2 protein - human pa
91	7	1.4	484	2	A40988	54K arginine-rich nuc
92	7	1.4	488	2	S36570	E2 protein - human pa
93	7	1.4	488	2	S18156	globulin 1 - eastern
94	7	1.4	493	2	S36488	E2 protein - human pa
95	7	1.4	494	2	A48133	pre-mRNA splicing SRP
96	7	1.4	498	1	W2WLB8	E2 protein - human pa

Tue Aug 10 13:01:48 1999

US-08-287-669-19.rag

Page 17

Db	1	mmrqdrrsl	l	lernimmf	sshlk	vd	e	ilev	liak	q	vl	nsd	ng	dm	ins	cg	t	vr	e	kr	re	i	v	k	a	60																													
Qy	1	mmr	odrrs	ll	ernim	f	ss	h	l	k	v	d	e	i	l	e	v	l	i	a	k	q	v	l	n	s	d	n	g	d	m	i	n	s	c	g	t	v	r	e	k	r	e	i	v	k	a	60							
Db	61	vgtr	gd	va	f	d	a	r	st	g	h	e	g	l	a	e	v	l	e	p	l	a	r	s	v	d	s	n	a	v	e	f	e	c	p	m	s	p	a	s	h	r	r	s	r	a	l	s	p	a	120				
Qy	61	vqrr	gd	va	f	d	a	r	st	g	h	e	g	l	a	e	v	l	e	p	l	a	r	s	v	d	s	n	a	v	e	f	e	c	p	m	s	p	a	s	h	r	r	s	r	a	l	s	p	a	120				
Db	121	gyts	pr	tv	h	r	d	s	v	s	s	v	s	f	t	s	y	q	d	i	y	s	r	a	r	s	r	s	r	a	l	h	s	s	d	r	h	y	s	p	p	v	n	a	f	s	p	q	s	180					
Qy	121	gyts	pr	tv	h	r	d	s	v	s	s	v	s	f	t	s	y	q	d	i	y	s	r	a	r	s	r	s	r	a	l	h	s	s	d	r	h	y	s	p	p	v	n	a	f	s	p	q	s	180					
Db	181	anss	f	t	g	c	s	s	l	g	y	s	s	s	r	n	r	s	f	s	k	a	s	g	p	t	y	i	f	h	e	e	d	m	n	f	v	d	a	p	t	i	s	r	v	f	e	k	t	m	y	n	f	s	240
Qy	181	anss	f	t	g	c	s	s	l	g	y	s	s	s	r	n	r	s	f	s	k	a	s	g	p	t	y	i	f	h	e	e	d	m	n	f	v	d	a	p	t	i	s	r	v	f	e	k	t	m	y	n	f	s	240
Db	241	prgm	c	x	i	n	n	e	h	e	f	e	m	p	t	r	n	g	t	k	a	d	n	i	n	l	n	f	r	c	m	y	t	v	i	c	k	n	i	t	g	r	m	l	t	i	r	d	f	a	k	300			
Qy	241	prgm	c	x	i	n	n	e	h	e	f	e	m	p	t	r	n	g	t	k	a	d	n	i	n	l	n	f	r	c	m	y	t	v	i	c	k	n	i	t	g	r	m	l	t	i	r	d	f	a	k	300			
Db	301	hesh	g	d	s	a	i	l	v	i	l	s	h	e	e	n	v	i	g	v	d	i	p	i	s	t	h	e	i	y	d	l	n	a	n	a	p	r	l	a	n	k	p	k	i	v	f	g	a	c	r	g	360		
Qy	301	hesh	g	d	s	a	i	l	v	i	l	s	h	e	e	n	v	i	g	v	d	i	p	i	s	t	h	e	i	y	d	l	n	a	n	a	p	r	l	a	n	k	p	k	i	v	f	g	a	c	r	g	360		
Db	361	errd	ng	f	p	l	d	s	v	d	g	v	p	a	f	i	r	r	g	w	n	r	d	p	i	n	f	i	g	c	v	r	p	q	v	q	w	r	k	p	s	q	a	d	i	l	i	r	y	a	420				
Qy	361	errd	ng	f	p	l	d	s	v	d	g	v	p	a	f	i	r	r	g	w	n	r	d	p	i	n	f	i	g	c	v	r	p	q	v	q	w	r	k	p	s	q	a	d	i	l	i	r	y	a	420				
Db	421	ttaq	y	s	w	n	s	a	r	g	s	w	i	q	a	v	c	e	v	f	s	t	h	a	k	m	d	v	e	l	l	e	v	n	k	k	v	a	466																

Search completed: Tue Aug 10 11:46:04 1999
Job time : 117 secs.

Db 421 ttaqyvswnsargswfigavcevfsthakmdvvealltevnkva 466
|||||
QY 421 TTAQYVSWRNSARGSWFIGAVCEVFSTHAKMDVVEALLTEVNKKVA 466

RESULT 14

ID R45303 standard; Protein: 503 AA.

AC R45303;

DT 01-JUL-1994 (first entry)

DE Ced-3 mutant G316.

KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
KW interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
KW inflammatory response; nematode; diagnosis; myocardial infarction;
KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
KW pathogenic infection; hair loss; cancer; autoreactive antibody.
OS Caenorhabditis elegans.

FH Key Location/Qualifiers

FT misc_difference 316

FT /label= G316

PN WO9325694-A.

PD 23-DEC-1993.

PF 14-JUN-1993; U05705.

PR 12-JUN-1992; US-897788.

PR 20-NOV-1992; US-984182.

PA (MASI) MASSACHUSETTS INST TECHNOLOGY.

PI Horvitz HR, Shaham S, Yuan J;

PI WPI; 94-007551/01.

PT Agents which affect activity of cell death genes - used to

PT develop drugs for treating conditions characterised by cell death

PT or proliferation

PS Claim 99; ; 132pp; English.

CC The sequences given in R45276-328 indicate the positions of possible
CC mutations of the C.elegans cell death gene, ced-3. These mutations
CC occur at conserved amino acid residues of the Ced-3 protein.

CC Fragments of the amino terminal of these proteins act as inhibitors of
CC Ced-3. The ced-3 gene has considerable similarity to human interleukin
CC -1beta convertase (ICE), which converts pro-interleukin-1beta to the
CC active cytokine and is involved in inflammatory response in humans.

CC The similarity between the two sequences suggests that inhibitors of
CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
CC proteins have an overall amino acid similarity of 28%. The ced-3

CC inhibitors may be used for identifying agents which affect the activity
CC of a gene belonging to the to the ced-3/ICE family of genes and for
CC diagnosis of diseases characterised by cell death. They can also be

CC used to develop drugs for treating conditions characterised by cell
CC deaths such as myocardial infarction, stroke, degenerative disease,
CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,

CC or drugs for reducing the proliferative capacity or size of a

CC population of cells such as cancerous cells, cells which produce

CC autoreactive antibodies, infected cells, hair follicle cells or cells

CC which are critical to the life of a parasite, pest or recombinant

CC organism. They may also be used in the diagnosis of inflammatory

CC disease.

CC Sequence 503 AA;

Query Match 92.4%; Score 465; DB 9; Length 503;

Best Local Similarity 99.8%; Pred. No. 0.00e+00;

Matches 465; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 mmrgdrrslernimmfshlkvdeilevliakvynsdngdmnscgtvrekrrivka 60
|||||

QY 1 MMRODRSLERNIMMFSSHLKVDLEILEVLIKQVLYNSDNGDMINSCGTVEKREIIVKA 60

Db 61 vgrgvdafadafdalrctgheglaevlleplarsvdsnavefecmpspashrralspa 120
|||||

QY 61 VORRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDSNAVEFECMPSPASHRRALSPA 120

Db 121 gytsptrvhrrdsvsvsftsdygdlysrarsrralhsdrhnyssppvnafpsqps 180
|||||

QY 121 GYTSPTRVHRDSSVSVSFTSYQDYISRARSRSRALHSSDRHNYSPPVNAFPSQPS 180

Db 181 ansftgcsslygsssrnsrfskasgptgyifhneedmfnvdaptisrvfdektmyrnfs 240
|||||

QY 181 ANSFTGCGSSLYGSSSRNSRFSKASGPTGYIFHEEDMNFVDAPTISRVFDEKTMYNFSS 240

Db 241 prgmclinnehfegmptnrngtkadkdnltlnfcmgytvickdnltgrgmlltirdfak 300
|||||

QY 241 PRGMCLINNHEFEGMPTNRNGTKADKDNLTNLFRCMGYTVICKDNLTGRGMLLTIRDFAK 300

Db 301 heshgdsailvlshxeenvligvddipisthelydllaanaaprlankpkivfgacrg 360
|||||

QY 301 HESHGDSAILVLISHXEENVIIGVDDIPISTHETIDLLAANAAPRLANKPKIVFVQACRG 360

Db 361 errdngfpvldsvdgvpafirrgwdnrdgplfnfigcvrpqvqvwrkpsqadiliry 420
|||||

QY 361 ERRDNGFPVLDSVDGVPAPFLRGGWDRDGPLFNLGCVRPQVQVWRKPSQADILIRYA 420

Db 421 ttaqyvswnsargswfigavcevfsthakmdvvealltevnkva 466
|||||

QY 421 TTAQYVSWRNSARGSWFIGAVCEVFSTHAKMDVVEALLTEVNKKVA 466

RESULT 15

ID R45279 standard; Protein: 503 AA.

AC R45279;

DT 01-JUL-1994 (first entry)

DE Ced-3 mutant L246.

KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
KW interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
KW inflammatory response; nematode; diagnosis; myocardial infarction;
KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
KW pathogenic infection; hair loss; cancer; autoreactive antibody.
OS Caenorhabditis elegans.

FH Key Location/Qualifiers

FT misc_difference 246

FT /label= L246

PN WO9325694-A.

PD 23-DEC-1993.

PF 14-JUN-1993; U05705.

PR 12-JUN-1992; US-897788.

PR 20-NOV-1992; US-984182.

PA (MASI) MASSACHUSETTS INST TECHNOLOGY.

PI Horvitz HR, Shaham S, Yuan J;

PI WPI; 94-007551/01.

PT Agents which affect activity of cell death genes - used to
PT develop drugs for treating conditions characterised by cell death

PT or proliferation

PS Claim 99; ; 132pp; English.

CC The sequences given in R45276-328 indicate the positions of possible
CC mutations of the C.elegans cell death gene, ced-3. These mutations
CC occur at conserved amino acid residues of the Ced-3 protein.

CC Fragments of the amino terminal of these proteins act as inhibitors of
CC Ced-3. The ced-3 gene has considerable similarity to human interleukin
CC -1beta convertase (ICE), which converts pro-interleukin-1beta to the

CC active cytokine and is involved in inflammatory response in humans.
CC The similarity between the two sequences suggests that inhibitors of
CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3

CC proteins have an overall amino acid similarity of 28%. The ced-3
CC inhibitors may be used for identifying agents which affect the activity
CC of a gene belonging to the to the ced-3/ICE family of genes and for
CC diagnosis of diseases characterised by cell death. They can also be

CC used to develop drugs for treating conditions characterised by cell
CC deaths such as myocardial infarction, stroke, degenerative disease,
CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,

CC or drugs for reducing the proliferative capacity or size of a

CC population of cells such as cancerous cells, cells which produce

CC autoreactive antibodies, infected cells, hair follicle cells or cells

CC which are critical to the life of a parasite, pest or recombinant

CC organism. They may also be used in the diagnosis of inflammatory

CC disease.

CC Sequence 503 AA;

Query Match 92.4%; Score 465; DB 9; Length 503;

Best Local Similarity 99.8%; Pred. No. 0.00e+00;

Matches 465; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

PT develop drugs for treating conditions characterised by cell death
 PT or proliferation
 PS Claim 99; ; 132pp; English.
 CC The sequences given in R45276-328 indicate the positions of possible
 CC mutations of the C.elegans cell death gene, ced-3. These mutations
 CC occur at conserved amino acid residues of the Ced-3 protein.
 CC Fragments of the amino terminal of these proteins act as inhibitors of
 CC Ced-3. The ced-3 gene has considerable similarity to human interleukin
 CC -1beta convertase (ICE), which converts pro-interleukin-1beta to the
 CC active cytokine and is involved in inflammatory response in humans.
 CC The similarity between the two sequences suggests that inhibitors of
 CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
 CC proteins have an overall amino acid similarity of 28%. The ced-3
 CC inhibitors may be used for identifying agents which affect the activity
 CC of a gene belonging to the to the ced-3/ICE family of genes and for
 CC diagnosis of diseases characterised by cell death. They can also be
 CC used to develop drugs for treating conditions characterised by cell
 CC deaths such as myocardial infarction, stroke, degenerative disease,
 CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
 CC or drugs for reducing the proliferative capacity or size of a
 CC population of cells such as cancerous cells, cells which produce
 CC autoreactive antibodies, infected cells, hair follicle cells or cells
 CC which are critical to the life of a parasite, pest or recombinant
 CC organism. They may also be used in the diagnosis of inflammatory
 CC disease.
 CC Sequence 503 AA;

Query Match 92.4%; Score 465; DB 9; Length 503;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 465; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 mmrgdrslrlnimmfshlkydeilleliakqvlnsdngdmnscgtvrekrrviak 60
 |||||||
 QY 1 MMRODRSLRLERNIMMFSSHLKVEILLEVLIAKQVLNSDNGDMINSCGTVREKREIVKA 60
 Db 61 vgrgvdafdafydairstgtgheglaevlleplarsvdsnavefecpmaspashrralspa 120
 |||||||
 QY 61 VQRGVDVAFDAFYDALRSTGHEGLAEVLEPLARSVDSNAVEFECPMSPASHRRALS SPA 120
 Db 121 gytsptrvhrdsvssvssftsdyqdiysrarsrsralhssdrhnyssppvnaifspgss 180
 |||||||
 QY 121 GYTSPTRVHRDSVSSVSSFTSYQDIYSRARSRSRALHSSDRHNYSSPPVNAIFSPQSS 180
 Db 181 ansstfgcsslgysssrnsrfskasgptqyifhnedmfvdaptisrvfdektmyrnfs 240
 |||||||
 QY 181 ANSSFTGCCSSLGYSRRNSRFSKASGPTQYIFHEEDMNFVDAPTISRVFDEKMYRNFS 240
 Db 241 prgmcllnnehfegmptnrgtkadkdnlnlfrcmgytvickdnltgrgmlltrdfak 300
 |||||||
 QY 241 PRGMCLLNNEHFEQMPTRNGTKADKDNLTNLFRCMGTYVICKDNLTGRGMLLTRDFAK 300
 Db 301 heshgdsailvilshgeenvilgyddipistheiydlinaaanprlankpkivfvacrg 360
 |||||||
 QY 301 HESHGDSAILVILSHGEENVITIGVDIPISTHEIYDLINAAANAPRLANKPKIVFVQACRG 360
 Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnlfgcvrpqvgvwrkksqadiliry 420
 |||||||
 QY 361 ERRDNGFPVLDSDVGPAPFLRRGWDNRDGPLFNFLGCVRPQVQVWRKKPSQADILIRYA 420
 Db 421 ttagyvswrnsargswifgavcefsthakmdvveillevnkkva 466
 |||||||
 QY 421 TTAQYVSWRNSARGSWIFGAVCEVSTHAKMDVVELLVEVNKKVA 466

RESULT 13
 ID R45294 standard; Protein; 503 AA.
 AC R45294;
 DT 01-JUL-1994 (first entry)
 DE Ced-3 mutant M291.
 KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
 KW interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
 KW inflammatory response; nematode; diagnosis; myocardial infarction;
 KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;

KW pathogenic infection; hair loss; cancer; autoreactive antibody.
 OS Caenorhabditis elegans.
 FH Key Location/Qualifiers
 FT misc_difference 291
 FT /label= M291

PN W09325694-A.
 PD 23-DEC-1993.
 PF 14-JUN-1993; U05705.
 PR 12-JUN-1992; US-897788.
 PR 20-NOV-1992; US-984182.
 PA (MAST) MASSACHUSETTS INST TECHNOLOGY.
 PI Horvitz HR, Shaham S, Yuan J;
 DR WPI; 94-007551/01.

PT Agents which affect activity of cell death genes - used to
 PT develop drugs for treating conditions characterised by cell death
 PT or proliferation

PS Claim 99; ; 132pp; English.
 CC The sequences given in R45276-328 indicate the positions of possible
 CC mutations of the C.elegans cell death gene, ced-3. These mutations
 CC occur at conserved amino acid residues of the Ced-3 protein.
 CC Fragments of the amino terminal of these proteins act as inhibitors of
 CC Ced-3. The ced-3 gene has considerable similarity to human interleukin
 CC -1beta convertase (ICE), which converts pro-interleukin-1beta to the
 CC active cytokine and is involved in inflammatory response in humans.
 CC The similarity between the two sequences suggests that inhibitors of
 CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
 CC proteins have an overall amino acid similarity of 28%. The ced-3
 CC inhibitors may be used for identifying agents which affect the activity
 CC of a gene belonging to the to the ced-3/ICE family of genes and for
 CC diagnosis of diseases characterised by cell death. They can also be
 CC used to develop drugs for treating conditions characterised by cell
 CC deaths such as myocardial infarction, stroke, degenerative disease,
 CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
 CC or drugs for reducing the proliferative capacity or size of a
 CC population of cells such as cancerous cells, cells which produce
 CC autoreactive antibodies, infected cells, hair follicle cells or cells
 CC which are critical to the life of a parasite, pest or recombinant
 CC organism. They may also be used in the diagnosis of inflammatory
 CC disease.
 CC Sequence 503 AA;

Query Match 92.4%; Score 465; DB 9; Length 503;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 465; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 mmrgdrslrlnimmfshlkydeilleliakqvlnsdngdmnscgtvrekrrviak 60
 |||||||
 QY 1 MMRODRSLRLERNIMMFSSHLKVEILLEVLIAKQVLNSDNGDMINSCGTVREKREIVKA 60
 Db 61 vgrgvdafdafydairstgtgheglaevlleplarsvdsnavefecpmaspashrralspa 120
 |||||||
 QY 61 VQRGVDVAFDAFYDALRSTGHEGLAEVLEPLARSVDSNAVEFECPMSPASHRRALS SPA 120
 Db 121 gytsptrvhrdsvssvssftsdyqdiysrarsrsralhssdrhnyssppvnaifspgss 180
 |||||||
 QY 121 GYTSPTRVHRDSVSSVSSFTSYQDIYSRARSRSRALHSSDRHNYSSPPVNAIFSPQSS 180
 Db 181 ansstfgcsslgysssrnsrfskasgptqyifhnedmfvdaptisrvfdektmyrnfs 240
 |||||||
 QY 181 ANSSFTGCCSSLGYSRRNSRFSKASGPTQYIFHEEDMNFVDAPTISRVFDEKMYRNFS 240
 Db 241 prgmcllnnehfegmptnrgtkadkdnlnlfrcmgytvickdnltgrgmlltrdfak 300
 |||||||
 QY 241 PRGMCLLNNEHFEQMPTRNGTKADKDNLTNLFRCMGTYVICKDNLTGRGMLLTRDFAK 300
 Db 301 heshgdsailvilshgeenvilgyddipistheiydlinaaanprlankpkivfvacrg 360
 |||||||
 QY 301 HESHGDSAILVILSHGEENVITIGVDIPISTHEIYDLINAAANAPRLANKPKIVFVQACRG 360
 Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnlfgcvrpqvgvwrkksqadiliry 420
 |||||||
 QY 361 ERRDNGFPVLDSDVGPAPFLRRGWDNRDGPLFNFLGCVRPQVQVWRKKPSQADILIRYA 420

Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 465; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 mmrqrdrslleerlnimmfsshlkvdeilevliakqvlinsdngdmnscgtvrekrrrelvka 60
|||||
QY 1 MMRODRSLLEERNIMFSSHLKVDLEILEVLI AKOV LNSDNGDMINSCGTVREKRRREL IYKA 60

Db 61 vqrrgdvafafadylalrstgtgheglaeleplarsvdsnavefecpmpashrrsralspa 120
|||||
QY 61 VQRGDVAFADFADYDALRSTGHEGLAELEPLARSDSNAVEFECPPMSPASHRRSRALSPA 120

Db 121 gytsptrvhrdsvssvssftsytqdllysrarsrsralhsdrhnyssppvnaafpsqps 180
|||||
QY 121 GYTSPTRVHRDSVSSVSSFTSYQDIYSRAKRSRSRALHSSDRHNYSSPPVNAFPSPQPS 180

Db 181 anssftgcslgysssrnsfskasgptqylfheedmnfvdaptrisrvfdektmyrnfs 240
|||||
QY 181 ANSFTGCSSLGYSSSRNRSFSKASGPTQYIFHEEDMNFVDAPTISR VFDEKTM YRNFS 240

Db 241 prgmcllinnehfeqmptnrgtkadknltnlfrcmgytvickdnltrgmlltirfak 300
|||||
QY 241 PRGMCLLINNEHFEQMPTNRGTKADKDNLTNLFRCMGYTVICKDNLTRGMLLTIRDFAK 300

Db 301 heshgdsalvylshgeenvliigvdipistheiydlinaanaprlankpkivfvqacrg 360
|||||
QY 301 HESHGDSALVYLSHGEENVLIIGVDIPISTHEIYDLINAANAPRLANKPKIVFVQACRG 360

Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnflgcvrpgvqvwrkpkpsqadilirya 420
|||||
QY 361 ERRDNGFPVLDSDVDGPAPFLRRGWDNRDGPLFNFLGCVRPQVQVWRKKPSQADILIRYA 420

Db 421 ttaqyvswnrsargswfigavcevfsthakmdvvellevnkva 466
|||||
QY 421 TTAQYVSWNRSGSWFIQAVCEVFSTHAKMDVVELLEVNKVA 466

RESULT 11
ID R45319 standard; Protein: 503 AA.
AC R45319;
DT 01-JUL-1994 (first entry)
DE Ced-3 mutant D414.
KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
interleukin-lbeta convertase; ICE; pro-interleukin-lbeta; cytokine;
inflammatory response; nematode; diagnosis; myocardial infarction;
stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
KW pathogenic infection; hair loss; cancer; autoreactive antibody.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT misc_difference 414 /label= D414

PN WO9325694-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05705.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-984182.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007551/01.
PT Agents which affect activity of cell death genes - used to
develop drugs for treating conditions characterised by cell death
or proliferation
PT Claim 99; 132pp; English.
PS The sequences given in R45276-328 indicate the positions of possible
mutations of the C.elegans cell death gene, ced-3. These mutations
occur at conserved amino acid residues of the Ced-3 protein.
CC Fragments of the amino terminal of these proteins act as inhibitors of
Ced-3. The ced-3 gene has considerable similarity to human interleukin
-lbeta convertase (ICE), which converts pro-interleukin-lbeta to the
active cytokine and is involved in inflammatory response in humans.
CC The similarity between the two sequences suggests that inhibitors of
ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
proteins have an overall amino acid similarity of 28%. The ced-3
inhibitors may be used for identifying agents which affect the activity

CC of a gene belonging to the to the ced-3/ICE family of genes and for
CC diagnosis of diseases characterised by cell death. They can also be
CC used to develop drugs for treating conditions characterised by cell
CC deaths such as myocardial infarction, stroke, degenerative disease,
CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
CC or drugs for reducing the proliferative capacity or size of a
CC population of cells such as cancerous cells, cells which produce
CC autoreactive antibodies, infected cells, hair follicle cells or cells
CC which are critical to the life of a parasite, pest or recombinant
CC organism. They may also be used in the diagnosis of inflammatory
CC disease.
SQ Sequence 503 AA;

Query Match 92.4%; Score 465; DB 9; Length 503;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 465; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 mmrqrdrslleerlnimmfsshlkvdeilevliakqvlinsdngdmnscgtvrekrrrelvka 60
|||||
QY 1 MMRODRSLLEERNIMFSSHLKVDLEILEVLI AKOV LNSDNGDMINSCGTVREKRRREL IYKA 60

Db 61 vqrrgdvafafadylalrstgtgheglaeleplarsvdsnavefecpmpashrrsralspa 120
|||||
QY 61 VQRGDVAFADFADYDALRSTGHEGLAELEPLARSDSNAVEFECPPMSPASHRRSRALSPA 120

Db 121 gytsptrvhrdsvssvssftsytqdllysrarsrsralhsdrhnyssppvnaafpsqps 180
|||||
QY 121 GYTSPTRVHRDSVSSVSSFTSYQDIYSRAKRSRSRALHSSDRHNYSSPPVNAFPSPQPS 180

Db 181 anssftgcslgysssrnsfskasgptqylfheedmnfvdaptrisrvfdektmyrnfs 240
|||||
QY 181 ANSFTGCSSLGYSSSRNRSFSKASGPTQYIFHEEDMNFVDAPTISR VFDEKTM YRNFS 240

Db 241 prgmcllinnehfeqmptnrgtkadknltnlfrcmgytvickdnltrgmlltirfak 300
|||||
QY 241 PRGMCLLINNEHFEQMPTNRGTKADKDNLTNLFRCMGYTVICKDNLTRGMLLTIRDFAK 300

Db 301 heshgdsalvylshgeenvliigvdipistheiydlinaanaprlankpkivfvqacrg 360
|||||
QY 301 HESHGDSALVYLSHGEENVLIIGVDIPISTHEIYDLINAANAPRLANKPKIVFVQACRG 360

Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnflgcvrpgvqvwrkpkpsqadilirya 420
|||||
QY 361 ERRDNGFPVLDSDVDGPAPFLRRGWDNRDGPLFNFLGCVRPQVQVWRKKPSQADILIRYA 420

Db 421 ttaqyvswnrsargswfigavcevfsthakmdvvellevnkva 466
|||||
QY 421 TTAQYVSWNRSGSWFIQAVCEVFSTHAKMDVVELLEVNKVA 466

RESULT 12
ID R45281 standard; Protein: 503 AA.
AC R45281;
DT 01-JUL-1994 (first entry)
DE Ced-3 mutant I248.
KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
interleukin-lbeta convertase; ICE; pro-interleukin-lbeta; cytokine;
inflammatory response; nematode; diagnosis; myocardial infarction;
stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
KW pathogenic infection; hair loss; cancer; autoreactive antibody.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT misc_difference 248 /label= I248

PN WO9325694-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05705.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-984182.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007551/01.
PT Agents which affect activity of cell death genes - used to

|||||
QY 361 ERRDNGFPVLDSVDGVPALFRGWDNRDGPLFNFLGCVRPQVOQWRRKPSQADILIRYA 420
Db 421 ttaqyvswnsarsgswfigavcevfsthakmdvveiltvknkva 466
|||||
QY 421 TTAQYVSWRNSARSGSWFIOAVCEVFSTHAKMDVVELLTVNKKVA 466

RESULT 9
ID R45326 standard; Protein: 503 AA.
AC R45326;
DT 01-JUL-1994 (first entry)
DE Ccd-3 mutant L488.
KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
inflammatory response; nematode; diagnosis; myocardial infarction;
stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
pathogenic infection; hair loss; cancer; autoreactive antibody.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
TI misc_difference 488 /label= L488
PN WO9325694-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05705.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-984182.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007551/01.
PT Agents which affect activity of cell death genes - used to
develop drugs for treating conditions characterised by cell death
or proliferation
PS Claim 99; : 132pp; English.
CC The sequences given in R45276-328 indicate the positions of possible
mutations of the C.elegans cell death gene, ced-3. These mutations
occur at conserved amino acid residues of the Ced-3 protein.
CC Fragments of the amino terminal of these proteins act as inhibitors of
Ced-3. The ced-3 gene has considerable similarity to human interleukin
-1beta convertase (ICE), which converts pro-interleukin-1beta to the
active cytokine and is involved in inflammatory response in humans.
CC The similarity between the two sequences suggests that inhibitors of
ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
proteins have an overall amino acid similarity of 28%. The ced-3
inhibitors may be used for identifying agents which affect the activity
of a gene belonging to the to the ced-3/ICE family of genes and for
diagnosis of diseases characterised by cell death. They can also be
used to develop drugs for treating conditions characterised by cell
deaths such as myocardial infarction, stroke, degenerative disease,
traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
or drugs for reducing the proliferative capacity or size of a
population of cells such as cancerous cells, cells which produce
autoreactive antibodies, infected cells, hair follicle cells or cells
which are critical to the life of a parasite, pest or recombinant
organism. They may also be used in the diagnosis of inflammatory
disease.
SQ Sequence 503 AA;

Query Match 92.6%; Score 466; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 mmrgdrrslernimmfshlkvdeilevliakvynsdngdmnscgtvrekrrrelvka 60
|||||
QY 1 MMRQDRSLERNIMMFSHLLKVDEILEVLIKQVLYNSDNGDMINSCGTVREKREIVKA 60
Db 61 vqrrgdvafadafyalrstgtgheglavleplarsvdsnavefecpmshrrsralspa 120
|||||
QY 61 VQRRGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDSNAVEFECPMSPASHRRSRALESPA 120
Db 121 gytsptrvhrdsvssvftsytgdiysrarstrsralhsdrrhysppvnafpsqps 180
|||||
QY 121 GYTSPTRVHRDSVSSVSTSYQDIYSRARSTRSRALHSSDRHNYSPPVNAFPSQPS 180

Db 181 ansstgcsslgysssrnsrfskasgptqyifhnedmfvdaptisrvfdektmyrnfs 240
|||||
QY 181 ANSSFTGCSSLGYSSSRNSRFSKASGPTQYIFHEDMNFVDAPTISRVEDEKTMYNFSS 240
Db 241 prgmcllnehefegmptnrgtkadknltnlfrcmgytvickdnltrgmlltirdfak 300
|||||
QY 241 PRGMCLLINNEHEFEGMPTNRGTAKDKDNLTNLFRCMGYTVICKDNLTRGMULLTRDFAK 300
Db 301 heshgdsailvishgeenviigvddipistheiydlinaanaprlankpkivfgacrg 360
|||||
QY 301 HESHGDSAILVILSHGEENVIIIGVDDIPISTHEIYDLLNAANAPRLANKPKIVFOACRG 360
Db 361 errdngfpvlsvdgvpaflirgwdnrdgplfnflgcvrpqvgvwrkpsqadiliry 420
|||||
QY 361 ERRDNGFPVLDSVDGVPALFRGWDNRDGPLFNFLGCVRPQVOQWRRKPSQADILIRYA 420
Db 421 ttaqyvswnsarsgswfigavcevfsthakmdvveiltvknkva 466
|||||
QY 421 TTAQYVSWRNSARSGSWFIOAVCEVFSTHAKMDVVELLTVNKKVA 466

RESULT 10
ID R45282 standard; Protein: 503 AA.
AC R45282;
DT 01-JUL-1994 (first entry)
DE Ccd-3 mutant N250.
KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
inflammatory response; nematode; diagnosis; myocardial infarction;
stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
pathogenic infection; hair loss; cancer; autoreactive antibody.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
TI misc_difference 250 /label= N250
PN WO9325694-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05705.
PR 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-984182.
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
DR WPI; 94-007551/01.
PT Agents which affect activity of cell death genes - used to
develop drugs for treating conditions characterised by cell death
or proliferation
PS Claim 99; : 132pp; English.
CC The sequences given in R45276-328 indicate the positions of possible
mutations of the C.elegans cell death gene, ced-3. These mutations
occur at conserved amino acid residues of the Ced-3 protein.
CC Fragments of the amino terminal of these proteins act as inhibitors of
Ced-3. The ced-3 gene has considerable similarity to human interleukin
-1beta convertase (ICE), which converts pro-interleukin-1beta to the
active cytokine and is involved in inflammatory response in humans.
CC The similarity between the two sequences suggests that inhibitors of
ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3
proteins have an overall amino acid similarity of 28%. The ced-3
inhibitors may be used for identifying agents which affect the activity
of a gene belonging to the to the ced-3/ICE family of genes and for
diagnosis of diseases characterised by cell death. They can also be
used to develop drugs for treating conditions characterised by cell
deaths such as myocardial infarction, stroke, degenerative disease,
traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
or drugs for reducing the proliferative capacity or size of a
population of cells such as cancerous cells, cells which produce
autoreactive antibodies, infected cells, hair follicle cells or cells
which are critical to the life of a parasite, pest or recombinant
organism. They may also be used in the diagnosis of inflammatory
disease.
SQ Sequence 503 AA;

Query Match 92.4%; Score 465; DB 9; Length 503;


```

Query Match          94.6%; Score 476; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 28 evliakvlnsdngdmnscgtvrekrrvkvavgrgdvafdafalrstgheglav 87
|||||
QY 28 EVLIAKQVLNSDNGDMINSCGTVREKRRREIKAVQRGDVAFDAFYDALRSTGHEGLAEV 87
|||||

Db 88 leplarsvdsnavefecpmpashrrsralspagytsptvrhdsavssvstsyqdiys 147
|||||
QY 88 LEPLARSDVSNAVEFECMPSPASHRRSRALSPAGYTSPTVRHDSVSSVSTSYQDIYS 147
|||||

Db 148 rarsrsralhsdrhnyssppvnaafpsqssansstfgcsslgysssrnrsfskasgp 207
|||||
QY 148 RARSRSRSLALHSSDRHNYSSPPVNAFSPSQSSANSSTFGCSSLGYSSSRNRSFSKASGP 207
|||||

Db 208 tqyifheedmnfvdaptrisvfdektmyrnfsprgmclinnhefegmptlrgtkadkd 267
|||||
QY 208 TQYIFHEEDMNFVDAPTRISRVFDEKTMYNFSSPRGMCLINNHEFEGMPTLRNGTKADKD 267
|||||

Db 268 nltnlfrcmgytvickdnltrgmlltirdfakhesghdsailvlshgeenvliigvddi 327
|||||
QY 268 NLTNLFRCMGYTVICKDNLTRGMLLTIRDFAKHESHGDSAILVLISHGEENVLIIGVDDI 327
|||||

Db 328 pistheiydlinaanaprlankpkivfgacrgerrndgfpvldsvdgvpaflrrgwdnr 387
|||||
QY 328 PISTHEIYDLINANAPRLANKPKIVFGACRGERRNDGFPVLDSVDGVPAPFLRRGWDNR 387
|||||

Db 388 dgpplfnlgcvrpgvqvrwkksqadiliryattaqyvswnrsarsgswfigavcevfst 447
|||||
QY 388 DGPLFNFLGCVRPQVQVWRKKPSQADILIRYATTAQYVSWRNSARSGSWFIQAVCEVFST 447
|||||

Db 448 hakdmvvellevnkvacgfgtsgsnlkkmpemtsrlkkfyfwearnsav 503
|||||
QY 448 HAKDMVVELLEVNKVACGFGTSGSNILKKMPEMTSRLKKFYFWPEARNSAV 503
|||||

RESULT 6
ID R45327 standard; Protein; 503 AA.
AC R45327;
DT 01-JUL-1994 (first entry)
DE Ced-3 mutant Y493.
KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
KW interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
KW inflammatory response; nematode; diagnosis; myocardial infarction;
KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
KW pathogenic infection; hair loss; cancer; autoreactive antibody.
KW Caenorhabditis elegans.
FH Key Location/Qualifiers
FT misc_difference 493
FT /label= Y493
PN WO9325694-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05705.
PK 12-JUN-1992; US-897788.
PK 20-NOV-1992; US-984182.
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI Horvitz HR, Shaham S, Yuan J;
PI 24-007551/01.
PT Agents which affect activity of cell death genes - used to
PT develop drugs for treating conditions characterised by cell death
PT or proliferation
PS Claim 99; ; 132pp; English.
CC The sequences given in R45276-328 indicate the positions of possible
CC mutations of the C.elegans cell death gene, ced-3. These mutations
CC occur at conserved amino acid residues of the Ced-3 protein.
CC Fragments of the amino terminal of these proteins act as inhibitors of
CC Ced-3. The ced-3 gene has considerable similarity to human interleukin
CC -1beta convertase (ICE), which converts pro-interleukin-1beta to the
CC active cytokine and is involved in inflammatory response in humans.
CC The similarity between the two sequences suggests that inhibitors of
CC ced-3 may also act as inhibitors of ICE. Human ICE and nematode Ced-3

```

```

CC proteins have an overall amino acid similarity of 28%. The ced-3
CC inhibitors may be used for identifying agents which affect the activity
CC of a gene belonging to the to the ced-3/ICE family of genes and for
CC diagnosis of diseases characterised by cell death. They can also be
CC used to develop drugs for treating conditions characterised by cell
CC deaths such as myocardial infarction, stroke, degenerative disease,
CC traumatic brain injury, hypoxia, pathogenic infection, or hair loss,
CC or drugs for reducing the proliferative capacity or size of a
CC population of cells such as cancerous cells, cells which produce
CC autoreactive antibodies, infected cells, hair follicle cells or cells
CC which are critical to the life of a parasite, pest or recombinant
CC organism. They may also be used in the diagnosis of inflammatory
CC disease.
SQ Sequence 503 AA;

Query Match          92.6%; Score 466; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mmrqdrsllemnfmfshlkvdeilevliakvlnsdngdmnscgtvrekrrvika 60
|||||
QY 1 MMRQDRSLLENNFMFSSHLKVDEILEVLIKQVLNSDNGDMINSCGTVREKRRVKA 60
|||||

Db 61 vqrrgdvafdafalrstgheglavleplarsvdsnavefecpmpashrrsralspa 120
|||||
QY 61 VQRRGDVAFDAFYDALRSTGHEGLAEVLEPLARSDVSNAVEFECMPSPASHRRSRALSPA 120
|||||

Db 121 gytsptrvhrdsvssvstsyqdiysrarsrsralhsdrhnyssppvnaafpsqps 180
|||||
QY 121 GYTSPTRVHRDSVSSVSTSYQDIYSRARSRSRALHSSDRHNYSSPPVNAFSPSQPS 180
|||||

Db 181 anssftgcsslgysssrnrsfskasgptqyifheedmnfvdaptrisvfdektmyrnfs 240
|||||
QY 181 ANSSFTGCSLGYSSSRNRSFSKASGPTQYIFHEEDMNFVDAPTRISRVFDEKTMYNFSS 240
|||||

Db 241 prgmclinnhefegmptlrgtkadkdnltrcmgytvickdnltrgmlltirdfak 300
|||||
QY 241 PRGMCLINNHEFEGMPTLRNGTKADKDNLTNFRCMGYTVICKDNLTRGMLLTIRDFAK 300
|||||

Db 301 heshgdsailvlshgeenvliigvddipistheiydlinaanaprlankpkivfvqacrg 360
|||||
QY 301 HESHGDSAILVLISHGEENVIIIGVDDIPISTHEIYDLINANAPRLANKPKIVFVQACRG 360
|||||

Db 361 errdngfpvldsvdgvpaflrrgwdnrdgplfnlgcvrpgvqvrwkksqadiliry 420
|||||
QY 361 ERRDNGFPVLDSVDGVPAPFLRRGWDNRDGPLFNFLGCVRPQVQVWRKKPSQADILIRYA 420
|||||

Db 421 ttaqyvswnrsarsgswfigavcefsthakdmvvellevnkvka 466
|||||
QY 421 TTAQYVSWRNSARSGSWFIQAVCEVFSTHAKDMVVELLEVNKVKA 466
|||||

RESULT 7
ID R45262 standard; Protein; 503 AA.
AC R45262;
DT 01-JUL-1994 (first entry)
DE Ced-3.
KW C.elegans; cell death; gene; ced-3; inhibition; human; parasite;
KW interleukin-1beta convertase; ICE; pro-interleukin-1beta; cytokine;
KW inflammatory response; nematode; diagnosis; myocardial infarction;
KW stroke; degenerative disease; traumatic brain injury; hypoxia; pest;
KW pathogenic infection; hair loss; cancer; autoreactive antibody.
KW Caenorhabditis elegans.
FH Key Location/Qualifiers
FT peptide 1.372
FT /note= "Claim 3, inhibitor fragment"
FT peptide 1.149
FT /note= "Claim 3, inhibitor fragment"
PN WO9325694-A.
PD 23-DEC-1993.
PF 14-JUN-1993; U05705.
PF 12-JUN-1992; US-897788.
PR 20-NOV-1992; US-984182.

```